

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
20 February 2003 (20.02.2003)

PCT

(10) International Publication Number
WO 03/013227 A2

(51) International Patent Classification⁷: **A01H**

(21) International Application Number: PCT/US02/25805

(22) International Filing Date: 9 August 2002 (09.08.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/310,847 9 August 2001 (09.08.2001) US
60/336,049 19 November 2001 (19.11.2001) US
60/338,692 11 December 2001 (11.12.2001) US
10/171,468 14 June 2002 (14.06.2002) US

(71) Applicant (for all designated States except US):
MENDEL BIOTECHNOLOGY, INC. [US/US]; 21375
Cabot Boulevard, Hayward, CA 94945 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **RATCLIFFE, Oliver** [GB/US]; 814 East 21st Street, Oakland, CA 94606 (US). **RIECHMANN, Jose, Luis** [ES/US]; 518 S. El Molino Avenue, #308, Pasadena, CA 91101 (US). **ADAM, Luc, J.** [CA/US]; 25800 Industrial Boulevard, Apt. L403, Hayward, CA 94545 (US). **DUBELL, Arnold, T.** [US/US]; 14857 Wake Avenue, San Leandro, CA 94578 (US). **HEARD, Jacqueline, E.** [US/US]; 810 Guilford Avenue, San Mateo, CA 94402 (US). **PILGRIM, Marsha, L.** [US/US]; 1368 Patrick Henry Drive, Phoenixville, PA 19460 (US). **JIANG, Cai-Zhong** [US/US]; 34495 Heathrow Terrace, Fremont, CA 94555 (US). **REUBER, T., Lynne** [US/US]; 1115 S. Grant Street, San Mateo, CA

94402 (US). **CREELMAN, Robert, A.** [US/US]; 2801 Jennifer Drive, Castro Valley, CA 94546 (US). **PINEDA, Omaira** [CO/US]; 4060 9th Place, Vero Beach, FL 32960 (US). **YU, Guo-Liang** [US/US]; 242 Gravett Drive, Berkeley, CA 94705-1531 (US). **BROUN, Pierre, E.** [FR/US]; 921 Sunnybrac Boulevard, San Mateo, CA 94402 (US).

(74) Agents: **WARD, Michael, R.** et al.; Morrison & Foerster LLP, 425 Market Street, San Francisco, CA 94105-2482 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: YIELD-RELATED POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS

(57) Abstract: The invention relates to plant transcription factor polypeptides, polynucleotides that encode them, homologs from a variety of plant species, and methods of using the polynucleotides and polypeptides to produce transgenic plants having advantageous properties compared to a reference plant. Sequence information related to these polynucleotides and polypeptides can also be used in bioinformatic search methods and is also disclosed.

WO 03/013227 A2

YIELD-RELATED POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS

This application claims the benefit of US Provisional Application No. 60/310,847, filed August 9, 2001, US Provisional Application No. 60/336,049, filed December 5, 2001, US Provisional Application No. 60/338,692, filed December 11, 2001, and US Non-provisional Application No. 10/171,468, filed June 14, 2002, the entire contents of which are hereby incorporated by reference.

FIELD OF THE INVENTION

This invention relates to the field of plant biology. More particularly, the present invention pertains to compositions and methods for phenotypically modifying a plant.

INTRODUCTION

A plant's traits, such as its biochemical, developmental, or phenotypic characteristics, may be controlled through a number of cellular processes. One important way to manipulate that control is through transcription factors - proteins that influence the expression of a particular gene or sets of genes. Transformed and transgenic plants that comprise cells having altered levels of at least one selected transcription factor, for example, possess advantageous or desirable traits. Strategies for manipulating traits by altering a plant cell's transcription factor content can therefore result in plants and crops with commercially valuable properties. Applicants have identified polynucleotides encoding transcription factors, developed numerous transgenic plants using these polynucleotides, and have analyzed the plants for a variety of important traits. In so doing, applicants have identified important polynucleotide and polypeptide sequences for producing commercially valuable plants and crops as well as the methods for making them and using them. Other aspects and embodiments of the invention are described below and can be derived from the teachings of this disclosure as a whole.

BACKGROUND OF THE INVENTION

Transcription factors (TFs) can modulate gene expression, either increasing or decreasing (inducing or repressing) the rate of transcription. This modulation results in differential levels of gene expression at various developmental stages, in different

tissues and cell types, and in response to different exogenous (e.g., environmental) and endogenous stimuli throughout the life cycle of the organism.

Because transcription factors are key controlling elements of biological pathways, altering the expression levels of one or more transcription factors can change entire biological pathways in an organism. For example, manipulation of the levels of selected transcription factors may result in increased expression of economically useful proteins or metabolic chemicals in plants or to improve other agriculturally relevant characteristics. Conversely, blocked or reduced expression of a transcription factor may reduce biosynthesis of unwanted compounds or remove an undesirable trait. Therefore, manipulating transcription factor levels in a plant offers tremendous potential in agricultural biotechnology for modifying a plant's traits.

The present invention provides novel transcription factors useful for modifying a plant's phenotype in desirable ways.

SUMMARY OF THE INVENTION

In a first aspect, the invention relates to a recombinant polynucleotide comprising a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence encoding a polypeptide comprising a polypeptide sequence selected from those of the Sequence Listing, SEQ ID NOs:2 to 2N, where $N = 2-561$, or those listed in Table 4, or a complementary nucleotide sequence thereof; (b) a nucleotide sequence encoding a polypeptide comprising a variant of a polypeptide of (a) having one or more, or between 1 and about 5, or between 1 and about 10, or between 1 and about 30, conservative amino acid substitutions; (c) a nucleotide sequence comprising a sequence selected from those of SEQ ID NOs:1 to $(2N - 1)$, where $N = 2-561$, or those included in Table 4, or a complementary nucleotide sequence thereof; (d) a nucleotide sequence comprising silent substitutions in a nucleotide sequence of (c); (e) a nucleotide sequence which hybridizes under stringent conditions over substantially the entire length of a nucleotide sequence of one or more of: (a), (b), (c), or (d); (f) a nucleotide sequence comprising at least 10 or 15, or at least about 20, or at least about 30 consecutive nucleotides of a sequence of any of (a)-(e), or at least 10 or 15, or at least about 20, or at least about 30 consecutive nucleotides outside of a region encoding a conserved domain of any of (a)-(e); (g) a

nucleotide sequence comprising a subsequence or fragment of any of (a)-(f), which subsequence or fragment encodes a polypeptide having a biological activity that modifies a plant's characteristic, functions as a transcription factor, or alters the level of transcription of a gene or transgene in a cell; (h) a nucleotide sequence having at least 31% sequence identity to a nucleotide sequence of any of (a)-(g); (i) a nucleotide sequence having at least 60%, or at least 70 %, or at least 80 %, or at least 90 %, or at least 95 % sequence identity to a nucleotide sequence of any of (a)-(g) or a 10 or 15 nucleotide, or at least about 20, or at least about 30 nucleotide region of a sequence of (a)-(g) that is outside of a region encoding a conserved domain; (j) a nucleotide sequence that encodes a polypeptide having at least 31% sequence identity to a polypeptide listed in Table 4, or the Sequence Listing; (k) a nucleotide sequence which encodes a polypeptide having at least 60%, or at least 70 %, or at least 80%, or at least 90 %, or at least 95 % sequence identity to a polypeptide listed in Table 4, or the Sequence Listing; and (l) a nucleotide sequence that encodes a conserved domain of a polypeptide having at least 85%, or at least 90%, or at least 95%, or at least 98% sequence identity to a conserved domain of a polypeptide listed in Table 4, or the Sequence Listing. The recombinant polynucleotide may further comprise a constitutive, inducible, or tissue-specific promoter operably linked to the nucleotide sequence. The invention also relates to compositions comprising at least two of the above-described polynucleotides.

In a second aspect, the invention comprises an isolated or recombinant polypeptide comprising a subsequence of at least about 10, or at least about 15, or at least about 20, or at least about 30 contiguous amino acids encoded by the recombinant or isolated polynucleotide described above, or comprising a subsequence of at least about 8, or at least about 12, or at least about 15, or at least about 20, or at least about 30 contiguous amino acids outside a conserved domain.

In a third aspect, the invention comprises an isolated or recombinant polynucleotide that encodes a polypeptide that is a paralog of the isolated polypeptide described above. In one aspect, the invention is an paralog which, when expressed in *Arabidopsis*, modifies a trait of the *Arabidopsis* plant.

In a fourth aspect, the invention comprises an isolated or recombinant polynucleotide that encodes a polypeptide that is an ortholog of the isolated polypeptide described above. In one aspect, the invention is an ortholog which, when expressed in *Arabidopsis*, modifies a trait of the *Arabidopsis* plant.

In a fifth aspect, the invention comprises an isolated polypeptide that is a paralog of the isolated polypeptide described above. In one aspect, the invention is an paralog which, when expressed in *Arabidopsis*, modifies a trait of the *Arabidopsis* plant.

In a sixth aspect, the invention comprises an isolated polypeptide that is an ortholog of the isolated polypeptide described above. In one aspect, the invention is an ortholog which, when expressed in *Arabidopsis*, modifies a trait of the *Arabidopsis* plant.

The present invention also encompasses transcription factor variants. A preferred transcription factor variant is one having at least 40% amino acid sequence identity, a more preferred transcription factor variant is one having at least 50% amino acid sequence identity and a most preferred transcription factor variant is one having at least 65% amino acid sequence identity to the transcription factor amino acid sequence SEQ ID NOs:2 to 2N, where $N = 2-561$, and which contains at least one functional or structural characteristic of the transcription factor amino acid sequence. Sequences having lesser degrees of identity but comparable biological activity are considered to be equivalents.

In another aspect, the invention is a transgenic plant comprising one or more of the above-described isolated or recombinant polynucleotides. In yet another aspect, the invention is a plant with altered expression levels of a polynucleotide described above or a plant with altered expression or activity levels of an above-described polypeptide. Further, the invention is a plant lacking a nucleotide sequence encoding a polypeptide described above or substantially lacking a polypeptide described above. The plant may be any plant, including, but not limited to, *Arabidopsis*, mustard, soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower, alfalfa, sugarcane, turf, banana, blackberry, blueberry, strawberry,

raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, pumpkin, spinach, squash, sweet corn, tobacco, tomato, watermelon, rosaceous fruits, vegetable brassicas, and mint or other labiates. In yet another aspect, the inventions is an isolated plant material of a plant, including, but not limited to, plant tissue, fruit, seed, plant cell, embryo, protoplast, pollen, and the like. In yet another aspect, the invention is a transgenic plant tissue culture of regenerable cells, including, but not limited to, embryos, meristematic cells, microspores, protoplast, pollen, and the like.

In yet another aspect the invention is a transgenic plant comprising one or more of the above described polynucleotides wherein the encoded polypeptide is expressed and regulates transcription of a gene.

In a further aspect the invention provides a method of using the polynucleotide composition to breed a progeny plant from a transgenic plant including crossing plants, producing seeds from transgenic plants, and methods of breeding using transgenic plants, the method comprising transforming a plant with the polynucleotide composition to create a transgenic plant, crossing the transgenic plant with another plant, selecting seed, and growing the progeny plant from the seed.

In a further aspect, the invention provides a progeny plant derived from a parental plant wherein said progeny plant exhibits at least three fold greater messenger RNA levels than said parental plant, wherein the messenger RNA encodes a DNA-binding protein which is capable of binding to a DNA regulatory sequence and inducing expression of a plant trait gene, wherein the progeny plant is characterized by a change in the plant trait compared to said parental plant. In yet a further aspect, the progeny plant exhibits at least ten fold greater messenger RNA levels compared to said parental plant. In yet a further aspect, the progeny plant exhibits at least fifty fold greater messenger RNA levels compared to said parental plant.

In a further aspect, the invention relates to a cloning or expression vector comprising the isolated or recombinant polynucleotide described above or cells comprising the cloning or expression vector.

In yet a further aspect, the invention relates to a composition produced by incubating a polynucleotide of the invention with a nuclease, a restriction enzyme, a polymerase; a polymerase and a primer; a cloning vector, or with a cell.

Furthermore, the invention relates to a method for producing a plant having a modified trait. The method comprises altering the expression of an isolated or recombinant polynucleotide of the invention or altering the expression or activity of a polypeptide of the invention in a plant to produce a modified plant, and selecting the modified plant for a modified trait. In one aspect, the plant is a monocot plant. In another aspect, the plant is a dicot plant. In another aspect the recombinant polynucleotide is from a dicot plant and the plant is a monocot plant. In yet another aspect the recombinant polynucleotide is from a monocot plant and the plant is a dicot plant. In yet another aspect the recombinant polynucleotide is from a monocot plant and the plant is a monocot plant. In yet another aspect the recombinant polynucleotide is from a dicot plant and the plant is a dicot plant.

In another aspect, the invention is a transgenic plant comprising an isolated or recombinant polynucleotide encoding a polypeptide wherein the polypeptide is selected from the group consisting of SEQ ID NOs: 2 - 2N, where $N = 2-561$. In yet another aspect, the invention is a plant with altered expression levels of a polypeptide described above or a plant with altered expression or activity levels of an above-described polypeptide. Further, the invention is a plant lacking a polynucleotide sequence encoding a polypeptide described above or substantially lacking a polypeptide described above. The plant may be any plant, including, but not limited to, Arabidopsis, mustard, soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower, alfalfa, sugarcane, turf, banana, blackberry, blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, pumpkin, spinach, squash, sweet corn, tobacco, tomato, watermelon, rosaceous fruits, vegetable brassicas, and mint or other labiates. In yet another aspect, the inventions is an isolated plant material of a plant, including, but not limited to, plant tissue, fruit, seed, plant cell, embryo, protoplast, pollen, and the like. In yet another aspect, the

invention is a transgenic plant tissue culture of regenerable cells, including, but not limited to, embryos, meristematic cells, microspores, protoplast, pollen, and the like.

In another aspect, the invention relates to a method of identifying a factor that is modulated by or interacts with a polypeptide encoded by a polynucleotide of the invention. The method comprises expressing a polypeptide encoded by the polynucleotide in a plant; and identifying at least one factor that is modulated by or interacts with the polypeptide. In one embodiment the method for identifying modulating or interacting factors is by detecting binding by the polypeptide to a promoter sequence, or by detecting interactions between an additional protein and the polypeptide in a yeast two hybrid system, or by detecting expression of a factor by hybridization to a microarray, subtractive hybridization, or differential display.

In yet another aspect, the invention is a method of identifying a molecule that modulates activity or expression of a polynucleotide or polypeptide of interest. The method comprises placing the molecule in contact with a plant comprising the polynucleotide or polypeptide encoded by the polynucleotide of the invention and monitoring one or more of the expression level of the polynucleotide in the plant, the expression level of the polypeptide in the plant, and modulation of an activity of the polypeptide in the plant.

In yet another aspect, the invention relates to an integrated system, computer or computer readable medium comprising one or more character strings corresponding to a polynucleotide of the invention, or to a polypeptide encoded by the polynucleotide. The integrated system, computer or computer readable medium may comprise a link between one or more sequence strings to a modified plant trait.

In yet another aspect, the invention is a method for identifying a sequence similar or homologous to one or more polynucleotides of the invention, or one or more polypeptides encoded by the polynucleotides. The method comprises providing a sequence database, and querying the sequence database with one or more target sequences corresponding to the one or more polynucleotides or to the one or more polypeptides to identify one or more sequence members of the database that display sequence similarity or homology to one or more of the one or more target sequences.

The method may further comprise of linking the one or more of the polynucleotides of the invention, or encoded polypeptides, to a modified plant phenotype.

BRIEF DESCRIPTION OF THE SEQUENCE LISTING, TABLES, AND FIGURE

The Sequence Listing provides exemplary polynucleotide and polypeptide sequences of the invention. The traits associated with the use of the sequences are included in the Examples.

Diskette1 is a read-only memory computer-readable diskette and contains a copy of the Sequence Listing in ASCII text format. The Sequence Listing is named "SEQLIST514442002041" and is 929 kilobytes in size. The copy of the Sequence Listing on the diskette is hereby incorporated by reference in its entirety.

Table 4 shows the polynucleotides and polypeptides identified by SEQ ID NO; Mendel Gene ID No.; conserved domain of the polypeptide; and if the polynucleotide was tested in a transgenic assay. The first column shows the polynucleotide SEQ ID NO; the second column shows the Mendel Gene ID No., GID; the third column shows the trait(s) resulting from the knock out or overexpression of the polynucleotide in the transgenic plant; the fourth column shows the category of the trait; the fifth column shows the transcription factor family to which the polynucleotide belongs; the sixth column ("Comment"), includes specific effects and utilities conferred by the polynucleotide of the first column; the seventh column shows the SEQ ID NO of the polypeptide encoded by the polynucleotide; and the eighth column shows the amino acid residue positions of the conserved domain in amino acid (AA) co-ordinates.

Table 5 lists a summary of orthologous and homologous sequences identified using BLAST (tblastx program). The first column shows the polynucleotide sequence identifier (SEQ ID NO), the second column shows the corresponding cDNA identifier (Gene ID), the third column shows the orthologous or homologous polynucleotide GenBank Accession Number (Test Sequence ID), the fourth column shows the

calculated probability value that the sequence identity is due to chance (Smallest Sum Probability), the fifth column shows the plant species from which the test sequence was isolated (Test Sequence Species), and the sixth column shows the orthologous or homologous test sequence GenBank annotation (Test Sequence GenBank Annotation).

Figure 1 shows a phylogenetic tree of related plant families adapted from Daly et al. (2001 *Plant Physiology* 127:1328-1333).

Detailed Description of Exemplary Embodiments

In an important aspect, the present invention relates to polynucleotides and polypeptides, e.g. for modifying phenotypes of plants. Throughout this disclosure, various information sources are referred to and/or are specifically incorporated. The information sources include scientific journal articles, patent documents, textbooks, and World Wide Web browser-inactive page addresses, for example. While the reference to these information sources clearly indicates that they can be used by one of skill in the art, applicants specifically incorporate each and every one of the information sources cited herein, in their entirety, whether or not a specific mention of "incorporation by reference" is noted. The contents and teachings of each and every one of the information sources can be relied on and used to make and use embodiments of the invention.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a plant" includes a plurality of such plants, and a reference to "a stress" is a reference to one or more stresses and equivalents thereof known to those skilled in the art, and so forth.

The polynucleotide sequences of the invention encode polypeptides that are members of well-known transcription factor families, including plant transcription factor families, as disclosed in Table 4. Generally, the transcription factors encoded by the present sequences are involved in cell differentiation and proliferation and the regulation of growth. Accordingly, one skilled in the art would recognize that by expressing the present sequences in a plant, one may change the expression of

autologous genes or induce the expression of introduced genes. By affecting the expression of similar autologous sequences in a plant that have the biological activity of the present sequences, or by introducing the present sequences into a plant, one may alter a plant's phenotype to one with improved traits. The sequences of the invention may also be used to transform a plant and introduce desirable traits not found in the wild-type cultivar or strain. Plants may then be selected for those that produce the most desirable degree of over- or underexpression of target genes of interest and coincident trait improvement.

The sequences of the present invention may be from any species, particularly plant species, in a naturally occurring form or from any source whether natural, synthetic, semi-synthetic or recombinant. The sequences of the invention may also include fragments of the present amino acid sequences. In this context, a "fragment" refers to a fragment of a polypeptide sequence which is at least 5 to about 15 amino acids in length, most preferably at least 14 amino acids, and which retain some biological activity of a transcription factor. Where "amino acid sequence" is recited to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

As one of ordinary skill in the art recognizes, transcription factors can be identified by the presence of a region or domain of structural similarity or identity to a specific consensus sequence or the presence of a specific consensus DNA-binding site or DNA-binding site motif (see, for example, Riechmann et al., (2000) *Science* 290: 2105-2110). The plant transcription factors may belong to one of the following transcription factor families: the AP2 (APETALA2) domain transcription factor family (Riechmann and Meyerowitz (1998) *Biol. Chem.* 379:633-646); the MYB transcription factor family (Martin and Paz-Ares, (1997) *Trends Genet.* 13:67-73); the MADS domain transcription factor family (Riechmann and Meyerowitz (1997) *Biol. Chem.* 378:1079-1101); the WRKY protein family (Ishiguro and Nakamura (1994) *Mol. Gen. Genet.* 244:563-571); the ankyrin-repeat protein family (Zhang et al. (1992) *Plant Cell* 4:1575-1588); the zinc finger protein (Z) family (Klug and Schwabe (1995) *FASEB J.* 9: 597-604); the homeobox (HB) protein family (Buerklin in *Guidebook to the Homeobox Genes*, Duboule (ed.) (1994) Oxford University Press);

the CAAT-element binding proteins (Forsburg and Guarente (1989) *Genes Dev.* 3:1166-1178); the squamosa promoter binding proteins (SPB) (Klein et al. (1996) *Mol. Gen. Genet.* 1996 250:7-16); the NAM protein family (Souer et al. (1996) *Cell* 85:159-170); the IAA/AUX proteins (Rouse et al. (1998) *Science* 279:1371-1373); the HLH/MYC protein family (Littlewood et al. (1994) *Prot. Profile* 1:639-709); the DNA-binding protein (DBP) family (Tucker et al. (1994) *EMBO J.* 13:2994-3002); the bZIP family of transcription factors (Foster et al. (1994) *FASEB J.* 8:192-200); the Box P-binding protein (the BPF-1) family (da Costa e Silva et al. (1993) *Plant J.* 4:125-135); the high mobility group (HMG) family (Bustin and Reeves (1996) *Prog. Nucl. Acids Res. Mol. Biol.* 54:35-100); the scarecrow (SCR) family (Di Laurenzio et al. (1996) *Cell* 86:423-433); the GF14 family (Wu et al. (1997) *Plant Physiol.* 114:1421-1431); the polycomb (PCOMB) family (Kennison (1995) *Annu. Rev. Genet.* 29:289-303); the teosinte branched (TEO) family (Luo et al. (1996) *Nature* 383:794-799); the ABI3 family (Giraudat et al. (1992) *Plant Cell* 4:1251-1261); the triple helix (TH) family (Dehesh et al. (1990) *Science* 250:1397-1399); the EIL family (Chao et al. (1997) *Cell* 89:1133-44); the AT-HOOK family (Reeves and Nissen (1990) *J. Biol. Chem.* 265:8573-8582); the S1FA family (Zhou et al. (1995) *Nucleic Acids Res.* 23:1165-1169); the bZIPT2 family (Lu and Ferl (1995) *Plant Physiol.* 109:723); the YABBY family (Bowman et al. (1999) *Development* 126:2387-96); the PAZ family (Bohmert et al. (1998) *EMBO J.* 17:170-80); a family of miscellaneous (MISC) transcription factors including the DPBF family (Kim et al. (1997) *Plant J.* 11:1237-1251) and the SPF1 family (Ishiguro and Nakamura (1994) *Mol. Gen. Genet.* 244:563-571); the golden (GLD) family (Hall et al. (1998) *Plant Cell* 10:925-936), the TUBBY family (Boggin et al. (1999) *Science* 286:2119-2125), the heat shock family (Wu C (1995) *Annu Rev Cell Dev Biol* 11:441-469), the ENBP family (Christiansen et al (1996) *Plant Mol Biol* 32:809-821), the RING-zinc family (Jensen et al. (1998) *FEBS letters* 436:283-287), the PDBP family (Janik et al *Virology.* (1989) 168:320-329), the PCF family (Cubas P, et al. *Plant J.* (1999) 18:215-22), the SRS (SHI-related) family (Fridborg et al *Plant Cell* (1999) 11:1019-1032), the CPP (cysteine-rich polycomb-like) family (Cvitanich et al *Proc. Natl. Acad. Sci. U S A.* (2000) 97:8163-8168), the ARF (auxin response factor) family (Ulmasov, et al. (1999) *Proc. Natl. Acad. Sci. USA* 96: 5844-5849), the SWI/SNF family (Collingwood et al *J. Mol. End.* 23:255-275), the ACBF family (Seguin et al (1997) *Plant Mol Biol.* 35:281-291), PCGL (CG-1 like) family (da Costa e Silva et al.

(1994) *Plant Mol Biol.* 25:921-924) the ARID family (Vazquez et al. (1999) *Development.* 126: 733-42), the Jumonji family, Balciunas et al (2000, *Trends Biochem Sci.* 25: 274-276), the bZIP-NIN family (Schauser et al (1999) *Nature* 402: 191-195), the E2F family Kaelin et al (1992) *Cell* 70: 351-364) and the GRF-like family (Knaap et al (2000) *Plant Physiol.* 122: 695-704). As indicated by any part of the list above and as known in the art, transcription factors have been sometimes categorized by class, family, and sub-family according to their structural content and consensus DNA-binding site motif, for example. Many of the classes and many of the families and sub-families are listed here. However, the inclusion of one sub-family and not another, or the inclusion of one family and not another, does not mean that the invention does not encompass polynucleotides or polypeptides of a certain family or sub-family. The list provided here is merely an example of the types of transcription factors and the knowledge available concerning the consensus sequences and consensus DNA-binding site motifs that help define them as known to those of skill in the art (each of the references noted above are specifically incorporated herein by reference). A transcription factor may include, but is not limited to, any polypeptide that can activate or repress transcription of a single gene or a number of genes. This polypeptide group includes, but is not limited to, DNA-binding proteins, DNA-binding protein binding proteins, protein kinases, protein phosphatases, GTP-binding proteins, and receptors, and the like.

In addition to methods for modifying a plant phenotype by employing one or more polynucleotides and polypeptides of the invention described herein, the polynucleotides and polypeptides of the invention have a variety of additional uses. These uses include their use in the recombinant production (i.e., expression) of proteins; as regulators of plant gene expression, as diagnostic probes for the presence of complementary or partially complementary nucleic acids (including for detection of natural coding nucleic acids); as substrates for further reactions, e.g., mutation reactions, PCR reactions, or the like; as substrates for cloning e.g., including digestion or ligation reactions; and for identifying exogenous or endogenous modulators of the transcription factors. A "polynucleotide" is a nucleic acid sequence comprising a plurality of polymerized nucleotides, e.g., at least about 15 consecutive polymerized nucleotides, optionally at least about 30 consecutive nucleotides, at least about 50 consecutive nucleotides. In many instances, a polynucleotide comprises a nucleotide

sequence encoding a polypeptide (or protein) or a domain or fragment thereof. Additionally, the polynucleotide may comprise a promoter, an intron, an enhancer region, a polyadenylation site, a translation initiation site, 5' or 3' untranslated regions, a reporter gene, a selectable marker, or the like. The polynucleotide can be single stranded or double stranded DNA or RNA. The polynucleotide optionally comprises modified bases or a modified backbone. The polynucleotide can be, e.g., genomic DNA or RNA, a transcript (such as an mRNA), a cDNA, a PCR product, a cloned DNA, a synthetic DNA or RNA, or the like. The polynucleotide can comprise a sequence in either sense or antisense orientations.

A "recombinant polynucleotide" is a polynucleotide that is not in its native state, e.g., the polynucleotide comprises a nucleotide sequence not found in nature, or the polynucleotide is in a context other than that in which it is naturally found, e.g., separated from nucleotide sequences with which it typically is in proximity in nature, or adjacent (or contiguous with) nucleotide sequences with which it typically is not in proximity. For example, the sequence at issue can be cloned into a vector, or otherwise recombined with one or more additional nucleic acid.

An "isolated polynucleotide" is a polynucleotide whether naturally occurring or recombinant, that is present outside the cell in which it is typically found in nature, whether purified or not. Optionally, an isolated polynucleotide is subject to one or more enrichment or purification procedures, e.g., cell lysis, extraction, centrifugation, precipitation, or the like.

A "polypeptide" is an amino acid sequence comprising a plurality of consecutive polymerized amino acid residues e.g., at least about 15 consecutive polymerized amino acid residues, optionally at least about 30 consecutive polymerized amino acid residues, at least about 50 consecutive polymerized amino acid residues. In many instances, a polypeptide comprises a polymerized amino acid residue sequence that is a transcription factor or a domain or portion or fragment thereof. Additionally, the polypeptide may comprise a localization domain, 2) an activation domain, 3) a repression domain, 4) an oligomerization domain or 5) a DNA-binding domain, or the like. The polypeptide optionally comprises modified

amino acid residues, naturally occurring amino acid residues not encoded by a codon, non-naturally occurring amino acid residues.

A "recombinant polypeptide" is a polypeptide produced by translation of a recombinant polynucleotide. A "synthetic polypeptide" is a polypeptide created by consecutive polymerization of isolated amino acid residues using methods well known in the art. An "isolated polypeptide," whether a naturally occurring or a recombinant polypeptide, is more enriched in (or out of) a cell than the polypeptide in its natural state in a wild type cell, e.g., more than about 5% enriched, more than about 10% enriched, or more than about 20%, or more than about 50%, or more, enriched, i.e., alternatively denoted: 105%, 110%, 120%, 150% or more, enriched relative to wild type standardized at 100%. Such an enrichment is not the result of a natural response of a wild type plant. Alternatively, or additionally, the isolated polypeptide is separated from other cellular components with which it is typically associated, e.g., by any of the various protein purification methods herein.

"Identity" or "similarity" refers to sequence similarity between two polynucleotide sequences or between two polypeptide sequences, with identity being a more strict comparison. The phrases "percent identity" and "% identity" refer to the percentage of sequence similarity found in a comparison of two or more polynucleotide sequences or two or more polypeptide sequences. Identity or similarity can be determined by comparing a position in each sequence that may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same nucleotide base or amino acid, then the molecules are identical at that position. A degree of similarity or identity between polynucleotide sequences is a function of the number of identical or matching nucleotides at positions shared by the polynucleotide sequences. A degree of identity of polypeptide sequences is a function of the number of identical amino acids at positions shared by the polypeptide sequences. A degree of homology or similarity of polypeptide sequences is a function of the number of amino acids, i.e., structurally related, at positions shared by the polypeptide sequences.

"Altered" nucleic acid sequences encoding polypeptide include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting

in a polynucleotide encoding a polypeptide with at least one functional characteristic of the polypeptide. Included within this definition are polymorphisms that may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding polypeptide, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding polypeptide. The encoded polypeptide protein may also be "altered", and may contain deletions, insertions, or substitutions of amino acid residues that produce a silent change and result in a functionally equivalent polypeptide. Deliberate amino acid substitutions may be made on the basis of similarity in residue side chain chemistry, including, but not limited to, polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological activity of polypeptide is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine. Alignments between different polypeptide sequences may be used to calculate "percentage sequence similarity".

The term "plant" includes whole plants, shoot vegetative organs/structures (e.g., leaves, stems and tubers), roots, flowers and floral organs/structures (e.g., bracts, sepals, petals, stamens, carpels, anthers and ovules), seed (including embryo, endosperm, and seed coat) and fruit (the mature ovary), plant tissue (e.g., vascular tissue, ground tissue, and the like) and cells (e.g., guard cells, egg cells, and the like), and progeny of same. The class of plants that can be used in the method of the invention is generally as broad as the class of higher and lower plants amenable to transformation techniques, including angiosperms (monocotyledonous and dicotyledonous plants), gymnosperms, ferns, horsetails, psilophytes, lycophytes, bryophytes, and multicellular algae. (See for example, Figure 1, adapted from Daly et al. 2001 *Plant Physiology* 127:1328-1333; and see also Tudge, C., The Variety of Life, Oxford University Press, New York, 2000, pp. 547-606.)

A "transgenic plant" refers to a plant that contains genetic material not found in a wild type plant of the same species, variety or cultivar. The genetic material may

include a transgene, an insertional mutagenesis event (such as by transposon or T-DNA insertional mutagenesis), an activation tagging sequence, a mutated sequence, a homologous recombination event or a sequence modified by chimeraplasty. Typically, the foreign genetic material has been introduced into the plant by human manipulation, but any method can be used as one of skill in the art recognizes.

A transgenic plant may contain an expression vector or cassette. The expression cassette typically comprises a polypeptide-encoding sequence operably linked (i.e., under regulatory control of) to appropriate inducible or constitutive regulatory sequences that allow for the expression of polypeptide. The expression cassette can be introduced into a plant by transformation or by breeding after transformation of a parent plant. A plant refers to a whole plant as well as to a plant part, such as seed, fruit, leaf, or root, plant tissue, plant cells or any other plant material, e.g., a plant explant, as well as to progeny thereof, and to *in vitro* systems that mimic biochemical or cellular components or processes in a cell.

"Ectopic expression or altered expression" in reference to a polynucleotide indicates that the pattern of expression in, e.g., a transgenic plant or plant tissue, is different from the expression pattern in a wild type plant or a reference plant of the same species. The pattern of expression may also be compared with a reference expression pattern in a wild type plant of the same species. For example, the polynucleotide or polypeptide is expressed in a cell or tissue type other than a cell or tissue type in which the sequence is expressed in the wild type plant, or by expression at a time other than at the time the sequence is expressed in the wild type plant, or by a response to different inducible agents, such as hormones or environmental signals, or at different expression levels (either higher or lower) compared with those found in a wild type plant. The term also refers to altered expression patterns that are produced by lowering the levels of expression to below the detection level or completely abolishing expression. The resulting expression pattern can be transient or stable, constitutive or inducible. In reference to a polypeptide, the term "ectopic expression or altered expression" further may relate to altered activity levels resulting from the interactions of the polypeptides with exogenous or endogenous modulators or from interactions with factors or as a result of the chemical modification of the polypeptides.

A "fragment" or "domain," with respect to a polypeptide, refers to a subsequence of the polypeptide. In some cases, the fragment or domain, is a subsequence of the polypeptide which performs at least one biological function of the intact polypeptide in substantially the same manner, or to a similar extent, as does the intact polypeptide. For example, a polypeptide fragment can comprise a recognizable structural motif or functional domain such as a DNA-binding site or domain that binds to a DNA promoter region, an activation domain, or a domain for protein-protein interactions. Fragments can vary in size from as few as 6 amino acids to the full length of the intact polypeptide, but are preferably at least about 30 amino acids in length and more preferably at least about 60 amino acids in length. In reference to a polynucleotide sequence, "a fragment" refers to any subsequence of a polynucleotide, typically, of at least about 15 consecutive nucleotides, preferably at least about 30 nucleotides, more preferably at least about 50 nucleotides, of any of the sequences provided herein.

The invention also encompasses production of DNA sequences that encode transcription factors and transcription factor derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding transcription factors or any fragment thereof.

A "conserved domain", with respect to a polypeptide, refers to a domain within a transcription factor family which exhibits a higher degree of sequence homology, such as at least 65% sequence identity including conservative substitutions, and preferably at least 80% sequence identity, and more preferably at least 85%, or at least about 86%, or at least about 87%, or at least about 88%, or at least about 90%, or at least about 95%, or at least about 98% amino acid residue sequence identity of a polypeptide of consecutive amino acid residues. A fragment or domain can be referred to as outside a consensus sequence or outside a consensus DNA-binding site that is known to exist or that exists for a particular transcription factor class, family, or sub-family. In this case, the fragment or domain will not include the exact amino acids of a consensus sequence or consensus DNA-binding

site of a transcription factor class, family or sub-family, or the exact amino acids of a particular transcription factor consensus sequence or consensus DNA-binding site. Furthermore, a particular fragment, region, or domain of a polypeptide, or a polynucleotide encoding a polypeptide, can be "outside a conserved domain" if all the amino acids of the fragment, region, or domain fall outside of a defined conserved domain(s) for a polypeptide or protein. The conserved domains for each of polypeptides of SEQ ID NOs:2 - 2N, where N = 2-561, are listed in Table 4 as described in Example VII. Also, many of the polypeptides of Table 4 have conserved domains specifically indicated by start and stop sites. A comparison of the regions of the polypeptides in SEQ ID NOs:2 - 2N, where N = 2-561, or of those in Table 4, allows one of skill in the art to identify conserved domain(s) for any of the polypeptides listed or referred to in this disclosure, including those in Table 4.

A "trait" refers to a physiological, morphological, biochemical, or physical characteristic of a plant or particular plant material or cell. In some instances, this characteristic is visible to the human eye, such as seed or plant size, or can be measured by biochemical techniques, such as detecting the protein, starch, or oil content of seed or leaves, or by observation of a metabolic or physiological process, e.g. by measuring uptake of carbon dioxide, or by the observation of the expression level of a gene or genes, e.g., by employing Northern analysis, RT-PCR, microarray gene expression assays, or reporter gene expression systems, or by agricultural observations such as stress tolerance, yield, or pathogen tolerance. Any technique can be used to measure the amount of, comparative level of, or difference in any selected chemical compound or macromolecule in the transgenic plants, however.

"Trait modification" refers to a detectable difference in a characteristic in a plant ectopically expressing a polynucleotide or polypeptide of the present invention relative to a plant not doing so, such as a wild type plant. In some cases, the trait modification can be evaluated quantitatively. For example, the trait modification can entail at least about a 2% increase or decrease in an observed trait (difference), at least a 5% difference, at least about a 10% difference, at least about a 20% difference, at least about a 30%, at least about a 50%, at least about a 70%, or at least about a 100%, or an even greater difference compared with a wild type plant. It is known that there can be a natural variation in the modified trait. Therefore, the trait modification

observed entails a change of the normal distribution of the trait in the plants compared with the distribution observed in wild type plant.

I. Traits Which May Be Modified

Trait modifications of particular interest include those to seed (such as embryo or endosperm), fruit, root, flower, leaf, stem, shoot, seedling or the like, including: enhanced tolerance to environmental conditions including freezing, chilling, heat, drought, water saturation, radiation and ozone; improved tolerance to microbial, fungal or viral diseases; improved tolerance to pest infestations, including nematodes, mollicutes, parasitic higher plants or the like; decreased herbicide sensitivity; improved tolerance of heavy metals or enhanced ability to take up heavy metals; improved growth under poor photoconditions (e.g., low light and/or short day length), or changes in expression levels of genes of interest. Other phenotype that can be modified relate to the production of plant metabolites, such as variations in the production of taxol, tocopherol, tocotrienol, sterols, phytosterols, vitamins, wax monomers, anti-oxidants, amino acids, lignins, cellulose, tannins, prenillipids (such as chlorophylls and carotenoids), glucosinolates, and terpenoids, enhanced or compositionally altered protein or oil production (especially in seeds), or modified sugar (insoluble or soluble) and/or starch composition. Physical plant characteristics that can be modified include cell development (such as the number of trichomes), fruit and seed size and number, yields of plant parts such as stems, leaves, inflorescences, and roots, the stability of the seeds during storage, characteristics of the seed pod (e.g., susceptibility to shattering), root hair length and quantity, internode distances, or the quality of seed coat. Plant growth characteristics that can be modified include growth rate, germination rate of seeds, vigor of plants and seedlings, leaf and flower senescence, male sterility, apomixis, flowering time, flower abscission, rate of nitrogen uptake, osmotic sensitivity to soluble sugar concentrations, biomass or transpiration characteristics, as well as plant architecture characteristics such as apical dominance, branching patterns, number of organs, organ identity, organ shape or size.

II. Transcription Factors Modify Expression Of Endogenous Genes

Expression of genes which encode transcription factors that modify expression of endogenous genes, polynucleotides, and proteins are well known in the art. In addition, transgenic plants comprising isolated polynucleotides encoding transcription

factors may also modify expression of endogenous genes, polynucleotides, and proteins. Examples include Peng et al. (1997, *Genes and Development* 11:3194-3205) and Peng et al. (1999, *Nature*, 400:256-261). In addition, many others have demonstrated that an *Arabidopsis* transcription factor expressed in an exogenous plant species elicits the same or very similar phenotypic response. See, for example, Fu et al. (2001, *Plant Cell* 13:1791-1802); Nandi et al. (2000, *Curr. Biol.* 10:215-218); Coupland (1995, *Nature* 377:482-483); and Weigel and Nilsson (1995, *Nature* 377:482-500).

In another example, Mandel et al. (1992, *Cell* 71:133-143) and Suzuki et al. (2001, *Plant J.* 28:409-418) teach that a transcription factor expressed in another plant species elicits the same or very similar phenotypic response of the endogenous sequence, as often predicted in earlier studies of *Arabidopsis* transcription factors in *Arabidopsis* (see Mandel et al., 1992, *supra*; Suzuki et al., 2001, *supra*).

Other examples include Müller et al. (2001, *Plant J.* 28:169-179); Kim et al. (2001, *Plant J.* 25:247-259); Kyoizuka and Shimamoto (2002, *Plant Cell Physiol.* 43:130-135); Boss and Thomas (2002, *Nature*, 416:847-850); He et al. (2000, *Transgenic Res.*, 9:223-227); and Robson et al. (2001, *Plant J.* 28:619-631).

In yet another example, Gilmour et al. (1998, *Plant J.* 16:433-442) teach an *Arabidopsis* AP2 transcription factor, CBF1, which, when overexpressed in transgenic plants, increases plant freezing tolerance. Jaglo et al (2001, *Plant Physiol.* 127:910-917) further identified sequences in *Brassica napus* which encode CBF-like genes and that transcripts for these genes accumulated rapidly in response to low temperature. Transcripts encoding CBF-like proteins were also found to accumulate rapidly in response to low temperature in wheat, as well as in tomato. An alignment of the CBF proteins from *Arabidopsis*, *B. napus*, wheat, rye, and tomato revealed the presence of conserved amino acid sequences, PKK/RPAGR_xKFxETRHP and DSAWR, that bracket the AP2/EREBP DNA binding domains of the proteins and distinguish them from other members of the AP2/EREBP protein family. (See Jaglo et al., *supra*.)

III. Polypeptides and Polynucleotides of the Invention

The present invention provides, among other things, transcription factors (TFs), and transcription factor homologue polypeptides, and isolated or recombinant polynucleotides encoding the polypeptides, or novel variant polypeptides or polynucleotides encoding novel variants of transcription factors derived from the specific sequences provided here. These polypeptides and polynucleotides may be employed to modify a plant's characteristic.

Exemplary polynucleotides encoding the polypeptides of the invention were identified in the *Arabidopsis thaliana* GenBank database using publicly available sequence analysis programs and parameters. Sequences initially identified were then further characterized to identify sequences comprising specified sequence strings corresponding to sequence motifs present in families of known transcription factors. In addition, further exemplary polynucleotides encoding the polypeptides of the invention were identified in the plant GenBank database using publicly available sequence analysis programs and parameters. Sequences initially identified were then further characterized to identify sequences comprising specified sequence strings corresponding to sequence motifs present in families of known transcription factors. Polynucleotide sequences meeting such criteria were confirmed as transcription factors.

Additional polynucleotides of the invention were identified by screening *Arabidopsis thaliana* and/or other plant cDNA libraries with probes corresponding to known transcription factors under low stringency hybridization conditions. Additional sequences, including full length coding sequences were subsequently recovered by the rapid amplification of cDNA ends (RACE) procedure, using a commercially available kit according to the manufacturer's instructions. Where necessary, multiple rounds of RACE are performed to isolate 5' and 3' ends. The full length cDNA was then recovered by a routine end-to-end polymerase chain reaction (PCR) using primers specific to the isolated 5' and 3' ends. Exemplary sequences are provided in the Sequence Listing.

The polynucleotides of the invention can be or were ectopically expressed in overexpressor or knockout plants and the changes in the characteristic(s) or trait(s) of

the plants observed. Therefore, the polynucleotides and polypeptides can be employed to improve the characteristics of plants.

The polynucleotides of the invention can be or were ectopically expressed in overexpressor plant cells and the changes in the expression levels of a number of genes, polynucleotides, and/or proteins of the plant cells observed. Therefore, the polynucleotides and polypeptides can be employed to change expression levels of a genes, polynucleotides, and/or proteins of plants.

IV. Producing Polypeptides

The polynucleotides of the invention include sequences that encode transcription factors and transcription factor homologue polypeptides and sequences complementary thereto, as well as unique fragments of coding sequence, or sequence complementary thereto. Such polynucleotides can be, e.g., DNA or RNA, e.g., mRNA, cRNA, synthetic RNA, genomic DNA, cDNA synthetic DNA, oligonucleotides, etc. The polynucleotides are either double-stranded or single-stranded, and include either, or both sense (i.e., coding) sequences and antisense (i.e., non-coding, complementary) sequences. The polynucleotides include the coding sequence of a transcription factor, or transcription factor homologue polypeptide, in isolation, in combination with additional coding sequences (e.g., a purification tag, a localization signal, as a fusion-protein, as a pre-protein, or the like), in combination with non-coding sequences (e.g., introns or inteins, regulatory elements such as promoters, enhancers, terminators, and the like), and/or in a vector or host environment in which the polynucleotide encoding a transcription factor or transcription factor homologue polypeptide is an endogenous or exogenous gene.

A variety of methods exist for producing the polynucleotides of the invention. Procedures for identifying and isolating DNA clones are well known to those of skill in the art, and are described in, e.g., Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volume 152 Academic Press, Inc., San Diego, CA ("Berger"); Sambrook et al., Molecular Cloning - A Laboratory Manual (2nd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989 ("Sambrook") and Current Protocols in Molecular Biology, F. M. Ausubel et al., eds.,

Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 2000) ("Ausubel").

Alternatively, polynucleotides of the invention, can be produced by a variety of in vitro amplification methods adapted to the present invention by appropriate selection of specific or degenerate primers. Examples of protocols sufficient to direct persons of skill through in vitro amplification methods, including the polymerase chain reaction (PCR) the ligase chain reaction (LCR), Qbeta-replicase amplification and other RNA polymerase mediated techniques (e.g., NASBA), e.g., for the production of the homologous nucleic acids of the invention are found in Berger (*supra*), Sambrook (*supra*), and Ausubel (*supra*), as well as Mullis et al., (1987) PCR Protocols A Guide to Methods and Applications (Innis et al. eds) Academic Press Inc. San Diego, CA (1990) (Innis). Improved methods for cloning in vitro amplified nucleic acids are described in Wallace et al., U.S. Pat. No. 5,426,039. Improved methods for amplifying large nucleic acids by PCR are summarized in Cheng et al. (1994) Nature 369: 684-685 and the references cited therein, in which PCR amplicons of up to 40kb are generated. One of skill will appreciate that essentially any RNA can be converted into a double stranded DNA suitable for restriction digestion, PCR expansion and sequencing using reverse transcriptase and a polymerase. *See, e.g., Ausubel, Sambrook and Berger, all supra.*

Alternatively, polynucleotides and oligonucleotides of the invention can be assembled from fragments produced by solid-phase synthesis methods. Typically, fragments of up to approximately 100 bases are individually synthesized and then enzymatically or chemically ligated to produce a desired sequence, e.g., a polynucleotide encoding all or part of a transcription factor. For example, chemical synthesis using the phosphoramidite method is described, e.g., by Beaucage et al. (1981) Tetrahedron Letters 22:1859-1869; and Matthes et al. (1984) EMBO J. 3:801-805. According to such methods, oligonucleotides are synthesized, purified, annealed to their complementary strand, ligated and then optionally cloned into suitable vectors. And if so desired, the polynucleotides and polypeptides of the invention can be custom ordered from any of a number of commercial suppliers.

V. Homologous Sequences

Sequences homologous, i.e., that share significant sequence identity or similarity, to those provided in the Sequence Listing, derived from *Arabidopsis thaliana* or from other plants of choice are also an aspect of the invention.

Homologous sequences can be derived from any plant including monocots and dicots and in particular agriculturally important plant species, including but not limited to, crops such as soybean, wheat, corn, potato, cotton, rice, rape, oilseed rape (including canola), sunflower, alfalfa, sugarcane and turf; or fruits and vegetables, such as banana, blackberry, blueberry, strawberry, and raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, pumpkin, spinach, squash, sweet corn, tobacco, tomato, watermelon, rosaceous fruits (such as apple, peach, pear, cherry and plum) and vegetable brassicas (such as broccoli, cabbage, cauliflower, Brussels sprouts, and kohlrabi). Other crops, fruits and vegetables whose phenotype can be changed include barley, rye, millet, sorghum, currant, avocado, citrus fruits such as oranges, lemons, grapefruit and tangerines, artichoke, cherries, nuts such as the walnut and peanut, endive, leek, roots, such as arrowroot, beet, cassava, turnip, radish, yam, and sweet potato, and beans. The homologous sequences may also be derived from woody species, such pine, poplar and eucalyptus, or mint or other labiates.

Orthologs And Paralogs

Several different methods are known by those of skill in the art for identifying and defining these functionally homologous sequences. Three general methods for defining paralog and ortholog are described; a paralog or ortholog or homolog may be identified by one or more of the methods described below.

Orthologs and paralog are evolutionarily related genes that have similar sequence and similar functions. Orthologs are structurally related genes in different species that are derived from a speciation event. Paralog are structurally related genes within a single species that are derived by a duplication event.

Within a single plant species, gene duplication may cause two copies of a particular gene, giving rise to two or more genes with similar sequence and similar function known as paralog. A paralog is therefore a similar gene with a similar function within the same species. Paralog typically cluster together or in the same

clade (a group of similar genes) when a gene family phylogeny is analyzed using programs such as CLUSTAL (Thompson et al. (1994) *Nucleic Acids Res.* 22:4673-4680; Higgins et al. (1996) *Methods Enzymol.* 266 383-402). Groups of similar genes can also be identified with pair-wise BLAST analysis (Feng and Doolittle (1987) *J. Mol. Evol.* 25:351-360). For example, a clade of very similar MADS domain transcription factors from *Arabidopsis* all share a common function in flowering time (Ratcliffe et al. (2001) *Plant Physiol.* 126:122-132), and a group of very similar AP2 domain transcription factors from *Arabidopsis* are involved in tolerance of plants to freezing (Gilmour et al. (1998) *Plant J.* 16:433-442). Analysis of groups of similar genes with similar function that fall within one clade can yield sub-sequences that are particular to the clade. These sub-sequences, known as consensus sequences, can not only be used to define the sequences within each clade, but define the functions of these genes; genes within a clade may contain paralogous or orthologous sequences that share the same function. (See also, for example, Mount, D.W. (2001) Bioinformatics: Sequence and Genome Analysis Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York page 543.)

Speciation, the production of new species from a parental species, can also give rise to two or more genes with similar sequence and similar function. These genes, termed orthologs, often have an identical function within their host plants and are often interchangeable between species without losing function. Because plants have common ancestors, many genes in any plant species will have a corresponding orthologous gene in another plant species. Once a phylogenetic tree for a gene family of one species has been constructed using a program such as CLUSTAL (Thompson et al. (1994) *Nucleic Acids Res.* 22:4673-4680; Higgins et al. (1996) *Methods Enzymol.* 266:383-402), potential orthologous sequences can be placed into the phylogenetic tree and its relationship to genes from the species of interest can be determined. Once the ortholog pair has been identified, the function of the test ortholog can be determined by determining the function of the reference ortholog.

Transcription factors that are homologous to the listed sequences will typically share at least about 30% amino acid sequence identity, or at least about 30% amino acid sequence identity outside of a known consensus sequence or consensus DNA-binding site. More closely related transcription factors can share at least about 50%,

about 60%, about 65%, about 70%, about 75% or about 80% or about 90% or about 95% or about 98% or more sequence identity with the listed sequences, or with the listed sequences but excluding or outside a known consensus sequence or consensus DNA-binding site, or with the listed sequences excluding one or all conserved domain. Factors that are most closely related to the listed sequences share, e.g., at least about 85%, about 90% or about 95% or more % sequence identity to the listed sequences, or to the listed sequences but excluding or outside a known consensus sequence or consensus DNA-binding site or outside one or all conserved domain. At the nucleotide level, the sequences will typically share at least about 40% nucleotide sequence identity, preferably at least about 50%, about 60%, about 70% or about 80% sequence identity, and more preferably about 85%, about 90%, about 95% or about 97% or more sequence identity to one or more of the listed sequences, or to a listed sequence but excluding or outside a known consensus sequence or consensus DNA-binding site, or outside one or all conserved domain. The degeneracy of the genetic code enables major variations in the nucleotide sequence of a polynucleotide while maintaining the amino acid sequence of the encoded protein. Conserved domains within a transcription factor family may exhibit a higher degree of sequence homology, such as at least 65% sequence identity including conservative substitutions, and preferably at least 80% sequence identity, and more preferably at least 85%, or at least about 86%, or at least about 87%, or at least about 88%, or at least about 90%, or at least about 95%, or at least about 98% sequence identity. Transcription factors that are homologous to the listed sequences should share at least 30%, or at least about 60%, or at least about 75%, or at least about 80%, or at least about 90%, or at least about 95% amino acid sequence identity over the entire length of the polypeptide or the homolog. In addition, transcription factors that are homologous to the listed sequences should share at least 30%, or at least about 60%, or at least about 75%, or at least about 80%, or at least about 90%, or at least about 95% amino acid sequence similarity over the entire length of the polypeptide or the homolog.

Percent identity can be determined electronically, e.g., by using the MEGALIGN program (DNASTAR, Inc. Madison, Wis.). The MEGALIGN program can create alignments between two or more sequences according to different methods, e.g., the clustal method. (See, e.g., Higgins, D. G. and P. M. Sharp (1988) Gene

73:237-244.) The clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. Other alignment algorithms or programs may be used, including FASTA, BLAST, or ENTREZ, FASTA and BLAST. These are available as a part of the GCG sequence analysis package (University of Wisconsin, Madison, Wis.), and can be used with or without default settings. ENTREZ is available through the National Center for Biotechnology Information. In one embodiment, the percent identity of two sequences can be determined by the GCG program with a gap weight of 1, e.g., each amino acid gap is weighted as if it were a single amino acid or nucleotide mismatch between the two sequences (see USPN 6,262,333).

Other techniques for alignment are described in *Methods in Enzymology*, vol. 266: *Computer Methods for Macromolecular Sequence Analysis* (1996), ed. Doolittle, Academic Press, Inc., San Diego, Calif., USA. Preferably, an alignment program that permits gaps in the sequence is utilized to align the sequences. The Smith-Waterman is one type of algorithm that permits gaps in sequence alignments. See *Methods Mol. Biol.* 70: 173-187 (1997). Also, the GAP program using the Needleman and Wunsch alignment method can be utilized to align sequences. An alternative search strategy uses MPSRCH software, which runs on a MASPAR computer. MPSRCH uses a Smith-Waterman algorithm to score sequences on a massively parallel computer. This approach improves ability to pick up distantly related matches, and is especially tolerant of small gaps and nucleotide sequence errors. Nucleic acid-encoded amino acid sequences can be used to search both protein and DNA databases.

The percentage similarity between two polypeptide sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no similarity between the two amino acid sequences are not included in determining percentage similarity. Percent identity between polynucleotide sequences can also be counted or calculated by other methods known in the art, e.g., the Jotun Hein method. (See, e.g., Hein, J. (1990) *Methods Enzymol.* 183:626-645.) Identity between sequences can also be determined by other methods

known in the art, e.g., by varying hybridization conditions (see US Patent Application No. 20010010913).

Thus, the invention provides methods for identifying a sequence similar or paralogous or orthologous or homologous to one or more polynucleotides as noted herein, or one or more target polypeptides encoded by the polynucleotides, or otherwise noted herein and may include linking or associating a given plant phenotype or gene function with a sequence. In the methods, a sequence database is provided (locally or across an inter or intra net) and a query is made against the sequence database using the relevant sequences herein and associated plant phenotypes or gene functions.

In addition, one or more polynucleotide sequences or one or more polypeptides encoded by the polynucleotide sequences may be used to search against a BLOCKS (Bairoch et al. (1997) *Nucleic Acids Res.* 25:217-221), PFAM, and other databases which contain previously identified and annotated motifs, sequences and gene functions. Methods that search for primary sequence patterns with secondary structure gap penalties (Smith et al. (1992) *Protein Engineering* 5:35-51) as well as algorithms such as Basic Local Alignment Search Tool (BLAST; Altschul, S. F. (1993) *J. Mol. Evol.* 36:290-300; Altschul et al. (1990) *supra*), BLOCKS (Henikoff, S. and Henikoff, G. J. (1991) *Nucleic Acids Research* 19:6565-6572), Hidden Markov Models (HMM; Eddy, S. R. (1996) *Cur. Opin. Str. Biol.* 6:361-365; Sonnhammer et al. (1997) *Proteins* 28:405-420), and the like, can be used to manipulate and analyze polynucleotide and polypeptide sequences encoded by polynucleotides. These databases, algorithms and other methods are well known in the art and are described in Ausubel et al. (1997; *Short Protocols in Molecular Biology*, John Wiley & Sons, New York N.Y., unit 7.7) and in Meyers, R. A. (1995; *Molecular Biology and Biotechnology*, Wiley VCH, New York N.Y., p 856-853).

Furthermore, methods using manual alignment of sequences similar or homologous to one or more polynucleotide sequences or one or more polypeptides encoded by the polynucleotide sequences may be used to identify regions of similarity and conserved domains. Such manual methods are well-known of those of skill in the art and can include, for example, comparisons of tertiary structure between a

polypeptide sequence encoded by a polynucleotide which comprises a known function with a polypeptide sequence encoded by a polynucleotide sequence which has a function not yet determined. Such examples of tertiary structure may comprise predicted alpha helices, beta-sheets, amphipathic helices, leucine zipper motifs, zinc finger motifs, proline-rich regions, cysteine repeat motifs, and the like.

VI. Identifying Polynucleotides or Nucleic Acids by Hybridization

Polynucleotides homologous to the sequences illustrated in the Sequence Listing and tables can be identified, e.g., by hybridization to each other under stringent or under highly stringent conditions. Single stranded polynucleotides hybridize when they associate based on a variety of well characterized physical-chemical forces, such as hydrogen bonding, solvent exclusion, base stacking and the like. The stringency of a hybridization reflects the degree of sequence identity of the nucleic acids involved, such that the higher the stringency, the more similar are the two polynucleotide strands. Stringency is influenced by a variety of factors, including temperature, salt concentration and composition, organic and non-organic additives, solvents, etc. present in both the hybridization and wash solutions and incubations (and number thereof), as described in more detail in the references cited above. Encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NOs: 860; 802; 240; 274; 558; 24; 1120; 44; 460; 286; 120; 130; 134; 698; 832; 580; 612; 48, and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G. M. and S. L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A. R. (1987) *Methods Enzymol.* 152:507-511.) Estimates of homology are provided by either DNA-DNA or DNA-RNA hybridization under conditions of stringency as is well understood by those skilled in the art (Hames and Higgins, Eds. (1985) *Nucleic Acid Hybridisation*, IRL Press, Oxford, U.K.). Stringency conditions can be adjusted to screen for moderately similar fragments, such as homologous sequences from distantly related organisms, to highly similar fragments, such as genes that duplicate functional enzymes from closely related organisms. Post-hybridization washes determine stringency conditions.

In addition to the nucleotide sequences listed in Tables 4 and 5, full length cDNA, orthologs, paralogs and homologs of the present nucleotide sequences may be

identified and isolated using well known methods. The cDNA libraries orthologs, paralogs and homologs of the present nucleotide sequences may be screened using hybridization methods to determine their utility as hybridization target or amplification probes.

An example of stringent hybridization conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or northern blot is about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Nucleic acid molecules that hybridize under stringent conditions will typically hybridize to a probe based on either the entire cDNA or selected portions, e.g., to a unique subsequence, of the cDNA under wash conditions of 0.2x SSC to 2.0 x SSC, 0.1% SDS at 50-65° C. For example, high stringency is about 0.2 x SSC, 0.1% SDS at 65° C. Ultra-high stringency will be the same conditions except the wash temperature is raised about 3 to about 5° C, and ultra-ultra-high stringency will be the same conditions except the wash temperature is raised about 6 to about 9° C. For identification of less closely related homologues washes can be performed at a lower temperature, e.g., 50° C. In general, stringency is increased by raising the wash temperature and/or decreasing the concentration of SSC, as known in the art.

In another example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and most preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, e.g., formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and most preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30° C, more preferably of at least about 37° C, and most preferably of at least about 42° C. Varying additional parameters, such as hybridization time, the concentration of detergent, e.g., sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the

art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred embodiment, hybridization will occur at 30° C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37° C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 µg/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42° C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50% formamide, and 200 µg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

The washing steps that follow hybridization can also vary in stringency. Wash stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include temperature of at least about 25° C, more preferably of at least about 42° C. Another preferred set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1% SDS at 65° C. The most preferred high stringency washes are of at least about 68° C. For example, in a preferred embodiment, wash steps will occur at 25° C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 42° C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a most preferred embodiment, the wash steps will occur at 68° C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional variations on these conditions will be readily apparent to those skilled in the art (see U.S. Patent Application No. 20010010913).

As another example, stringent conditions can be selected such that an oligonucleotide that is perfectly complementary to the coding oligonucleotide hybridizes to the coding oligonucleotide with at least about a 5-10x higher signal to noise ratio than the ratio for hybridization of the perfectly complementary oligonucleotide to a nucleic acid encoding a transcription factor known as of the filing

date of the application. Conditions can be selected such that a higher signal to noise ratio is observed in the particular assay which is used, e.g., about 15x, 25x, 35x, 50x or more. Accordingly, the subject nucleic acid hybridizes to the unique coding oligonucleotide with at least a 2x higher signal to noise ratio as compared to hybridization of the coding oligonucleotide to a nucleic acid encoding known polypeptide. Again, higher signal to noise ratios can be selected, e.g., about 5x, 10x, 25x, 35x, 50x or more. The particular signal will depend on the label used in the relevant assay, e.g., a fluorescent label, a colorimetric label, a radioactive label, or the like.

Alternatively, transcription factor homolog polypeptides can be obtained by screening an expression library using antibodies specific for one or more transcription factors. With the provision herein of the disclosed transcription factor, and transcription factor homologue nucleic acid sequences, the encoded polypeptide(s) can be expressed and purified in a heterologous expression system (e.g., *E. coli*) and used to raise antibodies (monoclonal or polyclonal) specific for the polypeptide(s) in question. Antibodies can also be raised against synthetic peptides derived from transcription factor, or transcription factor homologue, amino acid sequences. Methods of raising antibodies are well known in the art and are described in Harlow and Lane (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, New York. Such antibodies can then be used to screen an expression library produced from the plant from which it is desired to clone additional transcription factor homologues, using the methods described above. The selected cDNAs can be confirmed by sequencing and enzymatic activity.

VII. Sequence Variations

It will readily be appreciated by those of skill in the art, that any of a variety of polynucleotide sequences are capable of encoding the transcription factors and transcription factor homologue polypeptides of the invention. Due to the degeneracy of the genetic code, many different polynucleotides can encode identical and/or substantially similar polypeptides in addition to those sequences illustrated in the Sequence Listing. Nucleic acids having a sequence that differs from the sequences shown in the Sequence Listing, or complementary sequences, that encode functionally equivalent peptides (i.e., peptides having some degree of equivalent or similar

biological activity) but differ in sequence from the sequence shown in the sequence listing due to degeneracy in the genetic code, are also within the scope of the invention.

Altered polynucleotide sequences encoding polypeptides include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polynucleotide encoding a polypeptide with at least one functional characteristic of the instant polypeptides. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding the instant polypeptides, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding the instant polypeptides.

Allelic variant refers to any of two or more alternative forms of a gene occupying the same chromosomal locus. Allelic variation arises naturally through mutation, and may result in phenotypic polymorphism within populations. Gene mutations can be silent (i.e., no change in the encoded polypeptide) or may encode polypeptides having altered amino acid sequence. The term allelic variant is also used herein to denote a protein encoded by an allelic variant of a gene. Splice variant refers to alternative forms of RNA transcribed from a gene. Splice variation arises naturally through use of alternative splicing sites within a transcribed RNA molecule, or less commonly between separately transcribed RNA molecules, and may result in several mRNAs transcribed from the same gene. Splice variants may encode polypeptides having altered amino acid sequence. The term splice variant is also used herein to denote a protein encoded by a splice variant of an mRNA transcribed from a gene.

Those skilled in the art would recognize that the polypeptide sequence G681, SEQ ID NO: 580, represents a single transcription factor; allelic variation and alternative splicing may be expected to occur. Allelic variants of the polypeptide sequence of SEQ ID NO: 579 can be cloned by probing cDNA or genomic libraries from different individual organisms according to standard procedures. Allelic variants of the DNA sequence shown in SEQ ID NO: 579, including those containing silent mutations and those in which mutations result in amino acid sequence changes, are within the scope of the present invention, as are proteins which are allelic variants

of SEQ ID NO: 580. cDNAs generated from alternatively spliced mRNAs, which retain the properties of the transcription factor are included within the scope of the present invention, as are polypeptides encoded by such cDNAs and mRNAs. Allelic variants and splice variants of these sequences can be cloned by probing cDNA or genomic libraries from different individual organisms or tissues according to standard procedures known in the art (see USPN 6,388,064).

For example, Table 1 illustrates, e.g., that the codons AGC, AGT, TCA, TCC, TCG, and TCT all encode the same amino acid: serine. Accordingly, at each position in the sequence where there is a codon encoding serine, any of the above trinucleotide sequences can be used without altering the encoded polypeptide.

Table 1

Amino acid			Possible Codons							
Alanine	Ala	A	GCA	GCC	GCG	GCU				
Cysteine	Cys	C	TGC	TGT						
Aspartic acid	Asp	D	GAC	GAT						
Glutamic acid	Glu	E	GAA	GAG						
Phenylalanine	Phe	F	TTC	TTT						
Glycine	Gly	G	GGA	GGC	GGG	GGT				
Histidine	His	H	CAC	CAT						
Isoleucine	Ile	I	ATA	ATC	ATT					
Lysine	Lys	K	AAA	AAG						
Leucine	Leu	L	TTA	TTG	CTA	CTC	CTG	CTT		
Methionine	Met	M	ATG							
Asparagine	Asn	N	AAC	AAT						
Proline	Pro	P	CCA	CCC	CCG	CCT				
Glutamine	Gln	Q	CAA	CAG						
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGT		
Serine	Ser	S	AGC	AGT	TCA	TCC	TCG	TCT		
Threonine	Thr	T	ACA	ACC	ACG	ACT				
Valine	Val	V	GTA	GTC	GTG	GTT				
Tryptophan	Trp	W	TGG							
Tyrosine	Tyr	Y	TAC	TAT						

Sequence alterations that do not change the amino acid sequence encoded by the polynucleotide are termed "silent" variations. With the exception of the codons ATG and TGG, encoding methionine and tryptophan, respectively, any of the possible codons for the same amino acid can be substituted by a variety of techniques, e.g., site-directed mutagenesis, available in the art. Accordingly, any and all such variations of a sequence selected from the above table are a feature of the invention.

In addition to silent variations, other conservative variations that alter one, or a few amino acids in the encoded polypeptide, can be made without altering the function of the polypeptide, these conservative variants are, likewise, a feature of the invention.

For example, substitutions, deletions and insertions introduced into the sequences provided in the Sequence Listing are also envisioned by the invention. Such sequence modifications can be engineered into a sequence by site-directed mutagenesis (Wu (ed.) Meth. Enzymol. (1993) vol. 217, Academic Press) or the other methods noted below. Amino acid substitutions are typically of single residues; insertions usually will be on the order of about from 1 to 10 amino acid residues; and deletions will range about from 1 to 30 residues. In preferred embodiments, deletions or insertions are made in adjacent pairs, e.g., a deletion of two residues or insertion of two residues. Substitutions, deletions, insertions or any combination thereof can be combined to arrive at a sequence. The mutations that are made in the polynucleotide encoding the transcription factor should not place the sequence out of reading frame and should not create complementary regions that could produce secondary mRNA structure. Preferably, the polypeptide encoded by the DNA performs the desired function.

Conservative substitutions are those in which at least one residue in the amino acid sequence has been removed and a different residue inserted in its place. Such substitutions generally are made in accordance with the Table 2 when it is desired to maintain the activity of the protein. Table 2 shows amino acids which can be substituted for an amino acid in a protein and which are typically regarded as conservative substitutions.

Table 2

Residue	Conservative Substitutions
Ala	Ser
Arg	Lys
Asn	Gln; His
Asp	Glu
Gln	Asn
Cys	Ser
Glu	Asp
Gly	Pro
His	Asn; Gln
Ile	Leu, Val
Leu	Ile; Val
Lys	Arg; Gln
Met	Leu; Ile
Phe	Met; Leu; Tyr
Ser	Thr; Gly
Thr	Ser; Val
Trp	Tyr
Tyr	Trp; Phe
Val	Ile; Leu

Similar substitutions are those in which at least one residue in the amino acid sequence has been removed and a different residue inserted in its place. Such substitutions generally are made in accordance with the Table 3 when it is desired to maintain the activity of the protein. Table 3 shows amino acids which can be substituted for an amino acid in a protein and which are typically regarded as structural and functional substitutions. For example, a residue in column 1 of Table 3

may be substituted with residue in column 2; in addition, a residue in column 2 of Table 3 may be substituted with the residue of column 1.

Table 3

Residue	Similar Substitutions
Ala	Ser; Thr; Gly; Val; Leu; Ile
Arg	Lys; His; Gly
Asn	Gln; His; Gly; Ser; Thr
Asp	Glu; Ser; Thr
Gln	Asn; Ala
Cys	Ser; Gly
Glu	Asp
Gly	Pro; Arg
His	Asn; Gln; Tyr; Phe; Lys; Arg
Ile	Ala; Leu; Val; Gly; Met
Leu	Ala; Ile; Val; Gly; Met
Lys	Arg; His; Gln; Gly; Pro
Met	Leu; Ile; Phe
Phe	Met; Leu; Tyr; Trp; His; Val; Ala
Ser	Thr; Gly; Asp; Ala; Val; Ile; His
Thr	Ser; Val; Ala; Gly
Trp	Tyr; Phe; His
Tyr	Trp; Phe; His
Val	Ala; Ile; Leu; Gly; Thr; Ser; Glu

Substitutions that are less conservative than those in Table 2 can be selected by picking residues that differ more significantly in their effect on maintaining (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain. The substitutions which in general are

expected to produce the greatest changes in protein properties will be those in which (a) a hydrophilic residue, e.g., seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g., leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g., lysyl, arginyl, or histidyl, is substituted for (or by) an electronegative residue, e.g., glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g., phenylalanine, is substituted for (or by) one not having a side chain, e.g., glycine.

VIII. Further Modifying Sequences of the Invention – Mutation/Forced Evolution

In addition to generating silent or conservative substitutions as noted, above, the present invention optionally includes methods of modifying the sequences of the Sequence Listing. In the methods, nucleic acid or protein modification methods are used to alter the given sequences to produce new sequences and/or to chemically or enzymatically modify given sequences to change the properties of the nucleic acids or proteins.

Thus, in one embodiment, given nucleic acid sequences are modified, e.g., according to standard mutagenesis or artificial evolution methods to produce modified sequences. The modified sequences may be created using purified natural polynucleotides isolated from any organism or may be synthesized from purified compositions and chemicals using chemical means well known to those of skill in the art. For example, Ausubel, *supra*, provides additional details on mutagenesis methods. Artificial forced evolution methods are described, for example, by Stemmer (1994) Nature 370:389-391, Stemmer (1994) Proc. Natl. Acad. Sci. USA 91:10747-10751, and U.S. Patents 5,811,238, 5,837,500, and 6,242,568. Methods for engineering synthetic transcription factors and other polypeptides are described, for example, by Zhang et al. (2000) J. Biol. Chem. 275:33850-33860, Liu et al. (2001) J. Biol. Chem. 276:11323-11334, and Isalan et al. (2001) Nature Biotechnol. 19:656-660. Many other mutation and evolution methods are also available and expected to be within the skill of the practitioner.

Similarly, chemical or enzymatic alteration of expressed nucleic acids and polypeptides can be performed by standard methods. For example, sequence can be modified by addition of lipids, sugars, peptides, organic or inorganic compounds, by the inclusion of modified nucleotides or amino acids, or the like. For example, protein modification techniques are illustrated in Ausubel, *supra*. Further details on chemical and enzymatic modifications can be found herein. These modification methods can be used to modify any given sequence, or to modify any sequence produced by the various mutation and artificial evolution modification methods noted herein.

Accordingly, the invention provides for modification of any given nucleic acid by mutation, evolution, chemical or enzymatic modification, or other available methods, as well as for the products produced by practicing such methods, e.g., using the sequences herein as a starting substrate for the various modification approaches.

For example, optimized coding sequence containing codons preferred by a particular prokaryotic or eukaryotic host can be used e.g., to increase the rate of translation or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life, as compared with transcripts produced using a non-optimized sequence. Translation stop codons can also be modified to reflect host preference. For example, preferred stop codons for *Saccharomyces cerevisiae* and mammals are TAA and TGA, respectively. The preferred stop codon for monocotyledonous plants is TGA, whereas insects and *E. coli* prefer to use TAA as the stop codon.

The polynucleotide sequences of the present invention can also be engineered in order to alter a coding sequence for a variety of reasons, including but not limited to, alterations which modify the sequence to facilitate cloning, processing and/or expression of the gene product. For example, alterations are optionally introduced using techniques which are well known in the art, e.g., site-directed mutagenesis, to insert new restriction sites, to alter glycosylation patterns, to change codon preference, to introduce splice sites, etc.

Furthermore, a fragment or domain derived from any of the polypeptides of the invention can be combined with domains derived from other transcription factors or synthetic domains to modify the biological activity of a transcription factor. For instance, a DNA-binding domain derived from a transcription factor of the invention can be combined with the activation domain of another transcription factor or with a synthetic activation domain. A transcription activation domain assists in initiating transcription from a DNA-binding site. Examples include the transcription activation region of VP16 or GAL4 (Moore et al. (1998) Proc. Natl. Acad. Sci. USA 95: 376-381; and Aoyama et al. (1995) Plant Cell 7:1773-1785), peptides derived from bacterial sequences (Ma and Ptashne (1987) Cell 51: 113-119) and synthetic peptides (Giniger and Ptashne, (1987) Nature 330:670-672).

IX. Expression and Modification of Polypeptides

Typically, polynucleotide sequences of the invention are incorporated into recombinant DNA (or RNA) molecules that direct expression of polypeptides of the invention in appropriate host cells, transgenic plants, in vitro translation systems, or the like. Due to the inherent degeneracy of the genetic code, nucleic acid sequences which encode substantially the same or a functionally equivalent amino acid sequence can be substituted for any listed sequence to provide for cloning and expressing the relevant homologue.

X. Vectors, Promoters, and Expression Systems

The present invention includes recombinant constructs comprising one or more of the nucleic acid sequences herein. The constructs typically comprise a vector, such as a plasmid, a cosmid, a phage, a virus (e.g., a plant virus), a bacterial artificial chromosome (BAC), a yeast artificial chromosome (YAC), or the like, into which a nucleic acid sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available.

General texts that describe molecular biological techniques useful herein, including the use and production of vectors, promoters and many other relevant

topics, include Berger, Sambrook and Ausubel, *supra*. Any of the identified sequences can be incorporated into a cassette or vector, e.g., for expression in plants. A number of expression vectors suitable for stable transformation of plant cells or for the establishment of transgenic plants have been described including those described in Weissbach and Weissbach, (1989) Methods for Plant Molecular Biology, Academic Press, and Gelvin et al., (1990) Plant Molecular Biology Manual, Kluwer Academic Publishers. Specific examples include those derived from a Ti plasmid of *Agrobacterium tumefaciens*, as well as those disclosed by Herrera-Estrella et al. (1983) Nature 303: 209, Bevan (1984) Nucl Acid Res. 12: 8711-8721, Klee (1985) Bio/Technology 3: 637-642, for dicotyledonous plants.

Alternatively, non-Ti vectors can be used to transfer the DNA into monocotyledonous plants and cells by using free DNA delivery techniques. Such methods can involve, for example, the use of liposomes, electroporation, microprojectile bombardment, silicon carbide whiskers, and viruses. By using these methods transgenic plants such as wheat, rice (Christou (1991) Bio/Technology 9: 957-962) and corn (Gordon-Kamm (1990) Plant Cell 2: 603-618) can be produced. An immature embryo can also be a good target tissue for monocots for direct DNA delivery techniques by using the particle gun (Weeks et al. (1993) Plant Physiol 102: 1077-1084; Vasil (1993) Bio/Technology 10: 667-674; Wan and Lemeaux (1994) Plant Physiol 104: 37-48, and for *Agrobacterium*-mediated DNA transfer (Ishida et al. (1996) Nature Biotech 14: 745-750).

Typically, plant transformation vectors include one or more cloned plant coding sequence (genomic or cDNA) under the transcriptional control of 5' and 3' regulatory sequences and a dominant selectable marker. Such plant transformation vectors typically also contain a promoter (e.g., a regulatory region controlling inducible or constitutive, environmentally-or developmentally-regulated, or cell- or tissue-specific expression), a transcription initiation start site, an RNA processing signal (such as intron splice sites), a transcription termination site, and/or a polyadenylation signal.

Examples of constitutive plant promoters which can be useful for expressing the TF sequence include: the cauliflower mosaic virus (CaMV) 35S promoter, which

confers constitutive, high-level expression in most plant tissues (*see, e.g.,* Odell et al. (1985) Nature 313:810-812); the nopaline synthase promoter (An et al. (1988) Plant Physiol 88:547-552); and the octopine synthase promoter (Fromm et al. (1989) Plant Cell 1: 977-984).

A variety of plant gene promoters that regulate gene expression in response to environmental, hormonal, chemical, developmental signals, and in a tissue-active manner can be used for expression of a TF sequence in plants. Choice of a promoter is based largely on the phenotype of interest and is determined by such factors as tissue (e.g., seed, fruit, root, pollen, vascular tissue, flower, carpel, etc.), inducibility (e.g., in response to wounding, heat, cold, drought, light, pathogens, etc.), timing, developmental stage, and the like. Numerous known promoters have been characterized and can favorably be employed to promote expression of a polynucleotide of the invention in a transgenic plant or cell of interest. For example, tissue specific promoters include: seed-specific promoters (such as the napin, phaseolin or DC3 promoter described in US Pat. No. 5,773,697), fruit-specific promoters that are active during fruit ripening (such as the *dru 1* promoter (US Pat. No. 5,783,393), or the 2A11 promoter (US Pat. No. 4,943,674) and the tomato polygalacturonase promoter (Bird et al. (1988) Plant Mol Biol 11:651), root-specific promoters, such as those disclosed in US Patent Nos. 5,618,988, 5,837,848 and 5,905,186, pollen-active promoters such as PTA29, PTA26 and PTA13 (US Pat. No. 5,792,929), promoters active in vascular tissue (Ringli and Keller (1998) Plant Mol Biol 37:977-988), flower-specific (Kaiser et al. (1995) Plant Mol Biol 28:231-243), pollen (Baerson et al. (1994) Plant Mol Biol 26:1947-1959), carpels (Ohl et al. (1990) Plant Cell 2:837-848), pollen and ovules (Baerson et al. (1993) Plant Mol Biol 22:255-267), auxin-inducible promoters (such as that described in van der Kop et al. (1999) Plant Mol Biol 39:979-990 or Baumann et al. (1999) Plant Cell 11:323-334), cytokinin-inducible promoter (Guevara-Garcia (1998) Plant Mol Biol 38:743-753), promoters responsive to gibberellin (Shi et al. (1998) Plant Mol Biol 38:1053-1060, Willmott et al. (1998) 38:817-825) and the like. Additional promoters are those that elicit expression in response to heat (Ainley et al. (1993) Plant Mol Biol 22: 13-23), light (e.g., the pea *rbcS-3A* promoter, Kuhlemeier et al. (1989) Plant Cell 1:471, and the maize *rbcS* promoter, Schaffner and Sheen (1991) Plant Cell 3: 997); wounding (e.g., *wun1*, Siebertz et al. (1989) Plant Cell 1: 961); pathogens (such as the PR-1

promoter described in Buchel et al. (1999) Plant Mol. Biol. 40:387-396, and the PDF1.2 promoter described in Manners et al. (1998) Plant Mol. Biol. 38:1071-80), and chemicals such as methyl jasmonate or salicylic acid (Gatz et al. (1997) Plant Mol Biol 48: 89-108). In addition, the timing of the expression can be controlled by using promoters such as those acting at senescence (An and Amazon (1995) Science 270: 1986-1988); or late seed development (Odell et al. (1994) Plant Physiol 106:447-458).

Plant expression vectors can also include RNA processing signals that can be positioned within, upstream or downstream of the coding sequence. In addition, the expression vectors can include additional regulatory sequences from the 3'-untranslated region of plant genes, e.g., a 3' terminator region to increase mRNA stability of the mRNA, such as the PI-II terminator region of potato or the octopine or nopaline synthase 3' terminator regions.

Additional Expression Elements

Specific initiation signals can aid in efficient translation of coding sequences. These signals can include, e.g., the ATG initiation codon and adjacent sequences. In cases where a coding sequence, its initiation codon and upstream sequences are inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only coding sequence (e.g., a mature protein coding sequence), or a portion thereof, is inserted, exogenous transcriptional control signals including the ATG initiation codon can be separately provided. The initiation codon is provided in the correct reading frame to facilitate transcription. Exogenous transcriptional elements and initiation codons can be of various origins, both natural and synthetic. The efficiency of expression can be enhanced by the inclusion of enhancers appropriate to the cell system in use.

Expression Hosts

The present invention also relates to host cells which are transduced with vectors of the invention, and the production of polypeptides of the invention (including fragments thereof) by recombinant techniques. Host cells are genetically engineered (i.e., nucleic acids are introduced, e.g., transduced, transformed or transfected) with the vectors of this invention, which may be, for example, a cloning vector or an expression vector comprising the relevant nucleic acids herein. The

vector is optionally a plasmid, a viral particle, a phage, a naked nucleic acid, *etc.* The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants, or amplifying the relevant gene. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to those skilled in the art and in the references cited herein, including, Sambrook and Ausubel.

The host cell can be a eukaryotic cell, such as a yeast cell, or a plant cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Plant protoplasts are also suitable for some applications. For example, the DNA fragments are introduced into plant tissues, cultured plant cells or plant protoplasts by standard methods including electroporation (Fromm et al., (1985) Proc. Natl. Acad. Sci. USA 82, 5824, infection by viral vectors such as cauliflower mosaic virus (CaMV) (Hohn et al., (1982) Molecular Biology of Plant Tumors, (Academic Press, New York) pp. 549-560; US 4,407,956), high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface (Klein et al., (1987) Nature 327, 70-73), use of pollen as vector (WO 85/01856), or use of *Agrobacterium tumefaciens* or *A. rhizogenes* carrying a T-DNA plasmid in which DNA fragments are cloned. The T-DNA plasmid is transmitted to plant cells upon infection by *Agrobacterium tumefaciens*, and a portion is stably integrated into the plant genome (Horsch et al. (1984) Science 233:496-498; Fraley et al. (1983) Proc. Natl. Acad. Sci. USA 80, 4803).

The cell can include a nucleic acid of the invention which encodes a polypeptide, wherein the cells expresses a polypeptide of the invention. The cell can also include vector sequences, or the like. Furthermore, cells and transgenic plants that include any polypeptide or nucleic acid above or throughout this specification, e.g., produced by transduction of a vector of the invention, are an additional feature of the invention.

For long-term, high-yield production of recombinant proteins, stable expression can be used. Host cells transformed with a nucleotide sequence encoding a polypeptide of the invention are optionally cultured under conditions suitable for the

expression and recovery of the encoded protein from cell culture. The protein or fragment thereof produced by a recombinant cell may be secreted, membrane-bound, or contained intracellularly, depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides encoding mature proteins of the invention can be designed with signal sequences which direct secretion of the mature polypeptides through a prokaryotic or eukaryotic cell membrane.

XI. Modified Amino Acid Residues

Polypeptides of the invention may contain one or more modified amino acid residues. The presence of modified amino acids may be advantageous in, for example, increasing polypeptide half-life, reducing polypeptide antigenicity or toxicity, increasing polypeptide storage stability, or the like. Amino acid residue(s) are modified, for example, co-translationally or post-translationally during recombinant production or modified by synthetic or chemical means.

Non-limiting examples of a modified amino acid residue include incorporation or other use of acetylated amino acids, glycosylated amino acids, sulfated amino acids, prenylated (e.g., farnesylated, geranylgeranylated) amino acids, PEG modified (e.g., "PEGylated") amino acids, biotinylated amino acids, carboxylated amino acids, phosphorylated amino acids, etc. References adequate to guide one of skill in the modification of amino acid residues are replete throughout the literature.

The modified amino acid residues may prevent or increase affinity of the polypeptide for another molecule, including, but not limited to, polynucleotide, proteins, carbohydrates, lipids and lipid derivatives, and other organic or synthetic compounds.

XII. Identification of Additional Factors

A transcription factor provided by the present invention can also be used to identify additional endogenous or exogenous molecules that can affect a phenotype or trait of interest. On the one hand, such molecules include organic (small or large molecules) and/or inorganic compounds that affect expression of (i.e., regulate) a particular transcription factor. Alternatively, such molecules include endogenous

molecules that are acted upon either at a transcriptional level by a transcription factor of the invention to modify a phenotype as desired. For example, the transcription factors can be employed to identify one or more downstream gene with which is subject to a regulatory effect of the transcription factor. In one approach, a transcription factor or transcription factor homologue of the invention is expressed in a host cell, e.g., a transgenic plant cell, tissue or explant, and expression products, either RNA or protein, of likely or random targets are monitored, e.g., by hybridization to a microarray of nucleic acid probes corresponding to genes expressed in a tissue or cell type of interest, by two-dimensional gel electrophoresis of protein products, or by any other method known in the art for assessing expression of gene products at the level of RNA or protein. Alternatively, a transcription factor of the invention can be used to identify promoter sequences (i.e., binding sites) involved in the regulation of a downstream target. After identifying a promoter sequence, interactions between the transcription factor and the promoter sequence can be modified by changing specific nucleotides in the promoter sequence or specific amino acids in the transcription factor that interact with the promoter sequence to alter a plant trait. Typically, transcription factor DNA-binding sites are identified by gel shift assays. After identifying the promoter regions, the promoter region sequences can be employed in double-stranded DNA arrays to identify molecules that affect the interactions of the transcription factors with their promoters (Bulyk et al. (1999) Nature Biotechnology 17:573-577).

The identified transcription factors are also useful to identify proteins that modify the activity of the transcription factor. Such modification can occur by covalent modification, such as by phosphorylation, or by protein-protein (homo or heteropolymer) interactions. Any method suitable for detecting protein-protein interactions can be employed. Among the methods that can be employed are co-immunoprecipitation, cross-linking and co-purification through gradients or chromatographic columns, and the two-hybrid yeast system.

The two-hybrid system detects protein interactions in vivo and is described in Chien et al. ((1991), Proc. Natl. Acad. Sci. USA 88:9578-9582) and is commercially available from Clontech (Palo Alto, Calif.). In such a system, plasmids are constructed that encode two hybrid proteins: one consists of the DNA-binding domain

of a transcription activator protein fused to the TF polypeptide and the other consists of the transcription activator protein's activation domain fused to an unknown protein that is encoded by a cDNA that has been recombined into the plasmid as part of a cDNA library. The DNA-binding domain fusion plasmid and the cDNA library are transformed into a strain of the yeast *Saccharomyces cerevisiae* that contains a reporter gene (e.g., lacZ) whose regulatory region contains the transcription activator's binding site. Either hybrid protein alone cannot activate transcription of the reporter gene. Interaction of the two hybrid proteins reconstitutes the functional activator protein and results in expression of the reporter gene, which is detected by an assay for the reporter gene product. Then, the library plasmids responsible for reporter gene expression are isolated and sequenced to identify the proteins encoded by the library plasmids. After identifying proteins that interact with the transcription factors, assays for compounds that interfere with the TF protein-protein interactions can be preformed.

XIII. Identification of Modulators

In addition to the intracellular molecules described above, extracellular molecules that alter activity or expression of a transcription factor, either directly or indirectly, can be identified. For example, the methods can entail first placing a candidate molecule in contact with a plant or plant cell. The molecule can be introduced by topical administration, such as spraying or soaking of a plant, and then the molecule's effect on the expression or activity of the TF polypeptide or the expression of the polynucleotide monitored. Changes in the expression of the TF polypeptide can be monitored by use of polyclonal or monoclonal antibodies, gel electrophoresis or the like. Changes in the expression of the corresponding polynucleotide sequence can be detected by use of microarrays, Northern, quantitative PCR, or any other technique for monitoring changes in mRNA expression. These techniques are exemplified in Ausubel et al. (eds) Current Protocols in Molecular Biology, John Wiley & Sons (1998, and supplements through 2001). Such changes in the expression levels can be correlated with modified plant traits and thus identified molecules can be useful for soaking or spraying on fruit, vegetable and grain crops to modify traits in plants.

Essentially any available composition can be tested for modulatory activity of expression or activity of any nucleic acid or polypeptide herein. Thus, available libraries of compounds such as chemicals, polypeptides, nucleic acids and the like can be tested for modulatory activity. Often, potential modulator compounds can be dissolved in aqueous or organic (e.g., DMSO-based) solutions for easy delivery to the cell or plant of interest in which the activity of the modulator is to be tested. Optionally, the assays are designed to screen large modulator composition libraries by automating the assay steps and providing compounds from any convenient source to assays, which are typically run in parallel (e.g., in microtiter formats on microtiter plates in robotic assays).

In one embodiment, high throughput screening methods involve providing a combinatorial library containing a large number of potential compounds (potential modulator compounds). Such "combinatorial chemical libraries" are then screened in one or more assays, as described herein, to identify those library members (particular chemical species or subclasses) that display a desired characteristic activity. The compounds thus identified can serve as target compounds.

A combinatorial chemical library can be, e.g., a collection of diverse chemical compounds generated by chemical synthesis or biological synthesis. For example, a combinatorial chemical library such as a polypeptide library is formed by combining a set of chemical building blocks (e.g., in one example, amino acids) in every possible way for a given compound length (i.e., the number of amino acids in a polypeptide compound of a set length). Exemplary libraries include peptide libraries, nucleic acid libraries, antibody libraries (see, e.g., Vaughn et al. (1996) Nature Biotechnology, 14(3):309-314 and PCT/US96/10287), carbohydrate libraries (see, e.g., Liang et al. Science (1996) 274:1520-1522 and U.S. Patent 5,593,853), peptide nucleic acid libraries (see, e.g., U.S. Patent 5,539,083), and small organic molecule libraries (see, e.g., benzodiazepines, Baum C&EN Jan 18, page 33 (1993); isoprenoids, U.S. Patent 5,569,588; thiazolidinones and metathiazanones, U.S. Patent 5,549,974; pyrrolidines, U.S. Patents 5,525,735 and 5,519,134; morpholino compounds, U.S. Patent 5,506,337) and the like.

Preparation and screening of combinatorial or other libraries is well known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, peptide libraries (see, e.g., U.S. Patent 5,010,175; Furka, (1991) Int. J. Pept. Prot. Res. 37:487-493; and Houghton et al. (1991) Nature 354:84-88). Other chemistries for generating chemical diversity libraries can also be used.

In addition, as noted, compound screening equipment for high-throughput screening is generally available, e.g., using any of a number of well known robotic systems that have also been developed for solution phase chemistries useful in assay systems. These systems include automated workstations including an automated synthesis apparatus and robotic systems utilizing robotic arms. Any of the above devices are suitable for use with the present invention, e.g., for high-throughput screening of potential modulators. The nature and implementation of modifications to these devices (if any) so that they can operate as discussed herein will be apparent to persons skilled in the relevant art.

Indeed, entire high throughput screening systems are commercially available. These systems typically automate entire procedures including all sample and reagent pipetting, liquid dispensing, timed incubations, and final readings of the microplate in detector(s) appropriate for the assay. These configurable systems provide high throughput and rapid start up as well as a high degree of flexibility and customization. Similarly, microfluidic implementations of screening are also commercially available.

The manufacturers of such systems provide detailed protocols the various high throughput. Thus, for example, Zymark Corp. provides technical bulletins describing screening systems for detecting the modulation of gene transcription, ligand binding, and the like. The integrated systems herein, in addition to providing for sequence alignment and, optionally, synthesis of relevant nucleic acids, can include such screening apparatus to identify modulators that have an effect on one or more polynucleotides or polypeptides according to the present invention.

In some assays it is desirable to have positive controls to ensure that the components of the assays are working properly. At least two types of positive controls are appropriate. That is, known transcriptional activators or inhibitors can be

incubated with cells/plants/ etc. in one sample of the assay, and the resulting increase/decrease in transcription can be detected by measuring the resulting increase in RNA/ protein expression, etc., according to the methods herein. It will be appreciated that modulators can also be combined with transcriptional activators or inhibitors to find modulators that inhibit transcriptional activation or transcriptional repression. Either expression of the nucleic acids and proteins herein or any additional nucleic acids or proteins activated by the nucleic acids or proteins herein, or both, can be monitored.

In an embodiment, the invention provides a method for identifying compositions that modulate the activity or expression of a polynucleotide or polypeptide of the invention. For example, a test compound, whether a small or large molecule, is placed in contact with a cell, plant (or plant tissue or explant), or composition comprising the polynucleotide or polypeptide of interest and a resulting effect on the cell, plant, (or tissue or explant) or composition is evaluated by monitoring, either directly or indirectly, one or more of: expression level of the polynucleotide or polypeptide, activity (or modulation of the activity) of the polynucleotide or polypeptide. In some cases, an alteration in a plant phenotype can be detected following contact of a plant (or plant cell, or tissue or explant) with the putative modulator, e.g., by modulation of expression or activity of a polynucleotide or polypeptide of the invention. Modulation of expression or activity of a polynucleotide or polypeptide of the invention may also be caused by molecular elements in a signal transduction second messenger pathway and such modulation can affect similar elements in the same or another signal transduction second messenger pathway.

XIV. Subsequences

Also contemplated are uses of polynucleotides, also referred to herein as oligonucleotides, typically having at least 12 bases, preferably at least 15, more preferably at least 20, 30, or 50 bases, which hybridize under at least highly stringent (or ultra-high stringent or ultra-ultra-high stringent conditions) conditions to a polynucleotide sequence described above. The polynucleotides may be used as probes, primers, sense and antisense agents, and the like, according to methods as noted *supra*.

Subsequences of the polynucleotides of the invention, including polynucleotide fragments and oligonucleotides are useful as nucleic acid probes and primers. An oligonucleotide suitable for use as a probe or primer is at least about 15 nucleotides in length, more often at least about 18 nucleotides, often at least about 21 nucleotides, frequently at least about 30 nucleotides, or about 40 nucleotides, or more in length. A nucleic acid probe is useful in hybridization protocols, e.g., to identify additional polypeptide homologues of the invention, including protocols for microarray experiments. Primers can be annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand, and then extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods. See Sambrook and Ausubel, *supra*.

In addition, the invention includes an isolated or recombinant polypeptide including a subsequence of at least about 15 contiguous amino acids encoded by the recombinant or isolated polynucleotides of the invention. For example, such polypeptides, or domains or fragments thereof, can be used as immunogens, e.g., to produce antibodies specific for the polypeptide sequence, or as probes for detecting a sequence of interest. A subsequence can range in size from about 15 amino acids in length up to and including the full length of the polypeptide.

To be encompassed by the present invention, an expressed polypeptide which comprises such a polypeptide subsequence performs at least one biological function of the intact polypeptide in substantially the same manner, or to a similar extent, as does the intact polypeptide. For example, a polypeptide fragment can comprise a recognizable structural motif or functional domain such as a DNA binding domain that binds to a specific DNA promoter region, an activation domain or a domain for protein-protein interactions.

XV. Production of Transgenic Plants

Modification of Traits

The polynucleotides of the invention are favorably employed to produce transgenic plants with various traits, or characteristics, that have been modified in a desirable manner, e.g., to improve the seed characteristics of a plant. For example, alteration of expression levels or patterns (e.g., spatial or temporal expression patterns) of one or more of the transcription factors (or transcription factor homologues) of the invention, as compared with the levels of the same protein found in a wild type plant, can be used to modify a plant's traits. An illustrative example of trait modification, improved characteristics, by altering expression levels of a particular transcription factor is described further in the Examples and the Sequence Listing.

Arabidopsis as a model system

Arabidopsis thaliana is the object of rapidly growing attention as a model for genetics and metabolism in plants. *Arabidopsis* has a small genome, and well documented studies are available. It is easy to grow in large numbers and mutants defining important genetically controlled mechanisms are either available, or can readily be obtained. Various methods to introduce and express isolated homologous genes are available (see Koncz, et al., eds. *Methods in Arabidopsis Research*. et al. (1992), World Scientific, New Jersey, New Jersey, in "Preface"). Because of its small size, short life cycle, obligate autogamy and high fertility, *Arabidopsis* is also a choice organism for the isolation of mutants and studies in morphogenetic and development pathways, and control of these pathways by transcription factors (Koncz, *supra*, p. 72). A number of studies introducing transcription factors into *A. thaliana* have demonstrated the utility of this plant for understanding the mechanisms of gene regulation and trait alteration in plants. See, for example, Koncz, *supra*, and U.S. Patent Number 6,417,428).

Arabidopsis genes in transgenic plants.

Expression of genes which encode transcription factors modify expression of endogenous genes, polynucleotides, and proteins are well known in the art. In addition, transgenic plants comprising isolated polynucleotides encoding transcription factors may also modify expression of endogenous genes, polynucleotides, and

proteins. Examples include Peng et al. (1997, Genes and Development 11:3194-3205) and Peng et al. (1999, Nature, 400:256-261). In addition, many others have demonstrated that an *Arabidopsis* transcription factor expressed in an exogenous plant species elicits the same or very similar phenotypic response. See, for example, Fu et al. (2001, Plant Cell 13:1791-1802); Nandi et al. (2000, Curr. Biol. 10:215-218); Coupland (1995, Nature 377:482-483); and Weigel and Nilsson (1995, Nature 377:482-500).

Homologous genes introduced into transgenic plants.

Homologous genes that may be derived from any plant, or from any source whether natural, synthetic, semi-synthetic or recombinant, and that share significant sequence identity or similarity to those provided by the present invention, may be introduced into plants, for example, crop plants, to confer desirable or improved traits. Consequently, transgenic plants may be produced that comprise a recombinant expression vector or cassette with a promoter operably linked to one or more sequences homologous to presently disclosed sequences. The promoter may be, for example, a plant or viral promoter.

The invention thus provides for methods for preparing transgenic plants, and for modifying plant traits. These methods include introducing into a plant a recombinant expression vector or cassette comprising a functional promoter operably linked to one or more sequences homologous to presently disclosed sequences. Plants and kits for producing these plants that result from the application of these methods are also encompassed by the present invention.

The complete descriptions of the traits associated with each polynucleotide of the invention is fully disclosed in Table 4, Table 5, and Table 6.

Table 4

Polynucleotide SEQ ID NO:	GID No.	Trait	Category	Family	Comment	Polypeptide SEQ ID NO:	Conserved domains
1	G1275	Architecture; size	Dev and morph	WRKY	Reduced apical dominance; small plant	2	(113-169)
3	G1411	Architecture	Dev and morph	AP2	Loss of apical dominance	4	(87-154)
5	G1488	Architecture; light response; size; seed protein content	Dev and morph; seed biochemistry	GATA/Zn	Reduced apical dominance, shorter stems; constitutive photomorphogenesis; reduced size; altered seed protein content	6	(221-246)
7	G1499	Architecture; flower; morphology; other	Dev and morph	HLH/MYC	Altered plant architecture; altered floral organ identity and development; dark green color	8	(118-181)
9	G1543	Architecture; flower; morphology; other; seed oil	Dev and morph; seed biochemistry	HB	Altered plant architecture; altered carpel shape; dark green color; decreased seed oil	10	(135-195)
11	G1635	Architecture; morphology; other; fertility	Dev and morph	MYB-related	Reduced apical dominance; pale green, smaller plants; reduced fertility	12	(44-104)
13	G1794	Architecture; light response; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Altered plant architecture; constitutive photomorphogenesis; altered seed oil and protein content	14	(182-248)
15	G1839	Architecture; size	Dev and morph	AP2	Altered plant architecture; reduced size	16	(118-184)
17	G2108	Architecture	Dev and morph	AP2	Altered inflorescence structure	18	(18-85)
19	G2291	Architecture; flowering time	Dev and morph; flowering time	AP2	Altered plant architecture; late flowering	20	(TBD)
21	G2452	Architecture; leaf	Dev and morph	MYB-related	Reduced apical dominance; pale green color	22	(27-213)
23	G2509	Architecture; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Reduced apical dominance; altered seed oil and protein content	24	(89-156)
25	G390	Architecture	Dev and morph	HB	Altered shoot development	26	(18-81)
27	G391	Architecture	Dev and morph	HB	Altered shoot development	28	(25-85)
29	G438	Architecture; stem	Dev and morph	HB	Reduced branching; reduced lignin	30	(22-85)

Table 4

31	G47	Architecture; stem; flowering time; altered seed oil content	Dev and morph; flowering time; seed biochemistry	AP2	Altered architecture and inflorescence development, structure of vascular tissues; late flowering; altered seed oil content	32	(11-80)
33	G559	Architecture; fertility	Dev and morph	bZIP	Loss of apical dominance; reduced fertility	34	(203-264)
35	G568	Architecture; flowering time	Dev and morph	bZIP	Altered branching; late flowering	36	(215-265)
37	G580	Architecture; flower	Dev and morph	bZIP	Altered inflorescences; altered flower development	38	(162-218)
39	G615	Architecture; fertility	Dev and morph	TEO	Altered plant architecture; little or no pollen production, poor filament elongation	40	(88-147)
41	G732	Architecture; flower; seed oil and protein	Dev and morph; seed biochemistry	bZIP	Reduced apical dominance; abnormal flowers; altered seed oil and protein content	42	(31-91)
43	G988	Architecture; fertility; flower; stem; seed oil and protein content	Dev and morph; seed biochemistry	SCR	Reduced lateral branching; reduced fertility; enlarged floral organs, short pedicels; thicker stem, altered distribution of vascular bundles; altered seed oil and protein content	44	(178-195)
45	G1519	Embryo lethal	Dev and morph	RING/C3HC4	Embryo lethal	46	(327-364)
47	G374	Embryo lethal	Dev and morph	Z-ZPF	Embryo lethal	48	(35-67, 245-277)
49	G877	Embryo lethal	Dev and morph	WRKY	Embryo lethal	50	(272-328, 487-603)
51	G1000	Fertility; size; flower; stem	Dev and morph	MYB-(R1)R2R3	Reduced fertility; small plant; reduced or absent petals and sepals; reduced inflorescence, stem elongation	52	(14-117)
53	G1067	Fertility; leaf; size	Dev and morph	AT-hook	Reduced fertility; altered leaf shape; small plant	54	(86-93)
55	G1075	Fertility; flower; leaf; size	Dev and morph	AT-hook	Reduced fertility; reduced or absent petals, sepals and stamens; altered leaf shape; small plant	56	(78-85)
57	G1266	Fertility; size	Dev and morph	AP2	Reduced fertility; small plant	58	(79-147)
59	G1311	Fertility; size	Dev and morph	MYB-(R1)R2R3	Reduced fertility; small plant	60	(11-112)
61	G1321	Fertility; flower	Dev and morph	MYB-(R1)R2R3	Poor fertility; altered flower morphology	62	(4-106)
63	G1326	Fertility; flower; size	Dev and morph	MYB-(R1)R2R3	Reduced fertility; petals and sepals are smaller; small plant	64	(18-121)
65	G1367	Fertility; size	Dev and morph	AT-hook	Reduced fertility; reduced size	66	(179-201, 262-285, 298-319, 335-357)
67	G1386	Fertility; size; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Reduced fertility; reduced size; altered seed oil and protein content	68	(TBD)

Table 4

69	G1421	Fertility; size; seed oil content	Dev and morph; seed biochemistry	AP2	Reduced fertility; small plant; altered seed oil content	70	(74-151)
71	G1453	Fertility; morphology: other			Reduced fertility; altered inflorescence development	72	(13-160)
73	G1560	Fertility; flower; size	Dev and morph	NAC	Reduced fertility; altered flower development; reduced size	74	(62-151)
75	G1594	Fertility; leaf; seed	Dev and morph	HS	Reduced fertility; altered leaf shape and development; large pale seed	76	(343-308)
77	G1750	Fertility; size; seed oil content	Dev and morph; seed biochemistry	HB	Reduced fertility; reduced size; increased seed oil content	78	(107-173)
79	G1947	Fertility; flower; seed protein content	Dev and morph; seed biochemistry	AP2	Reduced fertility; extended period of flowering; altered seed protein content	80	(37-120)
81	G2011	Fertility; size; seed oil and protein content	Dev and morph; seed biochemistry	HS	Reduced fertility; reduced size; altered seed oil and protein content	82	(56-147)
83	G2094	Fertility; leaf; size	Dev and morph	GATA/Zn	Reduced fertility; altered leaf development; reduced size	84	(43-68)
85	G2113	Fertility; leaf; seed protein content	Dev and morph; seed biochemistry		Reduced fertility; long petioles, altered orientation; altered seed protein content	86	(TBD)
87	G2115	Fertility; size	Dev and morph	AP2	Reduced fertility; reduced size	88	(46-115)
89	G2130	Fertility; size; senescence	Dev and morph	AP2	Reduced fertility; reduced size; early senescence	90	(93-160)
91	G2147	Fertility; size	Dev and morph	HLH/MYC	Reduced fertility; reduced size	92	(160-234)
93	G2156	Fertility; size; seed protein content	Dev and morph; seed biochemistry	AT-hook	Reduced fertility; reduced size; altered seed protein content	94	(66-86)
95	G2294	Fertility; size	Dev and morph	AP2	Reduced fertility; reduced size	96	(32-102)
97	G2510	Fertility; size	Dev and morph	AP2	Reduced fertility; reduced size	98	(41-108)
99	G2893	Fertility; flower; size	Dev and morph	MYB-(R1)R2R3	Reduced fertility; altered flower development; reduced size	100	(19-120)
101	G340	Fertility; size	Dev and morph	Z-C3H	Reduced fertility, size	102	(37-154)
103	G39	Fertility; size	Dev and morph	AP2	Reduced fertility, small plant	104	(24-90)
105	G439	Fertility; size	Dev and morph	AP2	Reduced fertility, small plant	106	(110-177)
107	G470	Fertility	Dev and morph	ARF	Short stamen filaments	108	(61-393)

Table 4

109	G652	Fertility; seed; flower; size; seed oil content	Dev and morph; seed biochemistry	Z-CLDSH	Reduced fertility; irregular shaped seed; altered flower development; reduced size, slow growth; altered seed oil content	110	(28-49, 137-151, 182-196)
111	G671	Fertility; flower; leaf; size; stem	Dev and morph; biochemistry	MYB-(R1)R2R3	Reduced fertility; reduced petal abscission; altered leaf shape; small plant; altered inflorescence stem structure	112	(15-115)
113	G779	Fertility; flower	Dev and morph; biochemistry	HLH/MYC	Reduced fertility, homeotic transformations	114	(126-182)
115	G962	Fertility; size	Dev and morph; biochemistry	NAC	Reduced fertility; small plant	116	(53-175)
117	G977	Fertility; leaf; morphology; other; size	Dev and morph; biochemistry	AP2	Reduced fertility; altered leaf shape; dark green; small plant	118	(5-72)
119	G1063	Flower; leaf; inflorescence; seed oil and protein content	Dev and morph; seed biochemistry	HLH/MYC	Altered flower development, ectopic carpel tissue; altered leaf shape, dark green color; altered inflorescence development; altered seed oil and protein content	120	(131-182)
121	G1140	Flower	Dev and morph; biochemistry	MADS	Altered flower development	122	(2-57)
123	G1425	Flower	Dev and morph; biochemistry	NAC	Altered flower and inflorescence development	124	(20-173)
125	G1449	Flower	Dev and morph; biochemistry	IAA	Altered flower structure	126	(48-53, 74-107, 122-152)
127	G1897	Flower; leaf; seed protein content	Dev and morph; seed biochemistry	Z-Dof	Altered flower development; altered leaf development; altered seed protein content	128	(34-62)
129	G2143	Flower; leaf; inflorescence	Dev and morph; biochemistry	HLH/MYC	Altered flower development, ectopic carpel tissue; altered leaf shape, dark green color; altered inflorescence development	130	(128-179)
131	G2535	Flower; seed protein content	Dev and morph; seed biochemistry	NAC	Altered flower development; altered seed protein content	132	(11-114)
133	G2557	Flower; leaf	Dev and morph; biochemistry	HLH/MYC	Altered flower development, ectopic carpel tissue; altered leaf shape, dark green color	134	(278-328)
135	G259	Flower; leaf	Dev and morph; biochemistry	HS	Altered flower development; altered leaf development	136	(27-131)
137	G353	Flower; leaf; size; seed protein content	Dev and morph; seed biochemistry	Z-C2H2	Short pedicels, downward pointing siliques; altered leaf development; reduced size; altered seed protein content	138	(41-61, 84-104)
139	G354	Flower; light response; size	Dev and morph; biochemistry	Z-C2H2	Short pedicels, downward pointing siliques; constitutive morphogenesis; reduced size	140	(42-62, 88-109)
141	G638	Flower; morphology; other	Dev and morph; biochemistry	TH	Altered flower development; multiple developmental defects	142	(119-206)

Table 4

143	G869	Flower; morphology; other; seed oil	Dev and morph; seed biochemistry	AP2	Abnormal anther development; altered seed fatty acids	144	(109-177)
145	G1645	Inflorescence; leaf	Dev and morph	MYB-(R1)R2R3	Altered inflorescence structure; altered leaf development	146	(90-210)
147	G1038	Leaf	Dev and morph	GARP	Altered leaf shape	148	(198-247)
149	G1073	Leaf; size; flowering time	Dev and morph; flowering time	AT-hook	Serrated leaves; increased plant size; flowering appears to be slightly delayed	150	(33-42, 78-175)
151	G1146	Leaf	Dev and morph	PAZ	Altered leaf development	152	(886-896)
153	G1267	Leaf; size	Dev and morph	WRKY	Dark green shiny leaves; small plant	154	(70-127)
155	G1269	Leaf	Dev and morph	MYB-related	Long petioles, upturned leaves	156	(27-83)
157	G1452	Leaf; trichome; flowering time	Dev and morph; flowering time	NAC	Altered leaf shape, dark green color; reduced trichome density; late flowering	158	(30-177)
159	G1494	Leaf; size; light response; seed	Dev and morph	HLH/MYC	Pale green leaves, altered leaf shape; reduced size; long hypocotyls; large, pale seeds	160	(261-311)
161	G1548	Leaf	Dev and morph	HB	Altered leaf development	162	(17-77)
163	G1574	Leaf	Dev and morph	SWI/SNF	Altered leaf development	164	(28-350)
165	G1586	Leaf; size	Dev and morph	HB	Narrow leaves; small plants	166	(21-81)
167	G1786	Leaf; light response; size	Dev and morph	MYB-(R1)R2R3	Dark green, small leaves with short petioles; photomorphogenesis in the dark; small plant	168	(TBD)
169	G1792	Leaf; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Dark green, shiny leaves; altered seed oil and protein content	170	(17-85)
171	G1865	Leaf; seed oil and protein content	Dev and morph; seed biochemistry	GRF-like	Altered leaf development; altered seed oil and protein content	172	(124-149)
173	G1886	Leaf; size	Dev and morph	Z-Dof	Chlorotic patches in leaves; reduced size	174	(17-59)
175	G1933	Leaf; size; seed protein content	Dev and morph; seed biochemistry	WRKY	Altered leaf development; reduced size; altered seed protein content	176	(205-263, 344-404)
177	G2059	Leaf; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Smaller, curled leaves; altered seed oil, protein content	178	(184-254)
179	G2105	Leaf; seed	Dev and morph	TH	Alterations in leaf surface; large, pale seeds	180	(100-153)
181	G2117	Leaf; seed oil and protein content	Dev and morph; seed biochemistry	bZIP	Small, dark green leaves; altered seed oil and protein content	182	(46-106)

Table 4

183	G2124	Leaf; seed protein content	Dev and morph; seed biochemistry	TEO	Altered leaf development; altered seed protein content	184	(75-132)
185	G2140	Leaf; root	Dev and morph	HLH/MYC	Altered leaf development; short roots	186	(167-242)
187	G2144	Leaf; light response; size; seed oil content	Dev and morph; seed biochemistry	HLH/MYC	Pale green leaves, altered leaf shape; long hypocotyls; reduced size; altered seed oil content	188	(203-283)
189	G2431	Leaf	Dev and morph	GARP	Dark green leaves; reduced size	190	(38-88)
191	G2465	Morphology; other; leaf	Dev and morph	GARP	Slowed development; altered leaf color and shape	192	(219-269)
193	G2583	Leaf; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Glossy, shiny leaves; altered seed oil and protein content	194	(4-71)
195	G2724	Leaf	Dev and morph	MYB-(R1)R2R3	Dark green leaves	196	(7-113)
197	G377	Leaf; morphology; other	Dev and morph	RING/C3H2C3	Altered leaf development; slow growth	198	(85-128)
199	G428	Leaf	Dev and morph	HB	Altered leaf shape	200	(229-292)
201	G447	Leaf; morphology; other; size	Dev and morph	ARF	Dark green leaves; altered cotyledon shape; reduced size	202	(22-356)
203	G464	Leaf	Dev and morph	IAA	Altered leaf shape	204	(20-28, 71-82, 126-142, 187-224)
205	G557	Leaf; size	Dev and morph	bZIP	Dark green color; small plant	206	(90-150)
207	G577	Leaf	Dev and morph	BZIP2	Reduced size, increased anthocyanins	208	(TBD)
209	G674	Leaf; size	Dev and morph	MYB-(R1)R2R3	Dark green leaves, upwardly oriented; reduced size	210	(20-120)
211	G736	Leaf; flowering time	Dev and morph; flowering time	Z-Dof	Altered leaf shape; later flowering	212	(54-111)
213	G903	Leaf	Dev and morph	Z-C2H2	Altered leaf morphology	214	(68-92)
215	G917	Leaf; seed oil and protein content	Dev and morph; seed biochemistry	MADS	Altered leaf development; altered seed oil and protein content	216	(2-57)
217	G921	Leaf	Dev and morph	WRKY	Serrated leaves	218	(146-203)
219	G922	Leaf; size	Dev and morph	SCR	Altered development, dark green color; reduced size	220	(225-242)
221	G932	Leaf; size	Dev and morph	MYB-(R1)R2R3	Altered development, dark green color; reduced size	222	(12-118)
223	G599	Leaf; size	Dev and morph	DBP	Altered leaf shape; small plant	224	(187-219, 264-300)
225	G804	Leaf; size	Dev and morph	PCF	Altered leaf shape, small plant	226	(54-117)

Table 4

227	G1062	Light response; morphology; other; seed	Dev and morph	HLH/MYC	Constitutive photomorphogenesis; slow growth; altered seed shape	228	(308-359)
229	G1322	Light response; size	Dev and morph	MYB-(R1)R2R3	Photomorphogenesis in the dark; reduced size	230	(26-130)
231	G1331	Light response; morphology; other; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Constitutive photomorphogenesis; multiple developmental alterations; altered seed oil and protein content	232	(8-109)
233	G1521	Light response	Dev and morph	RING/C3HC4	Constitutive photomorphogenesis	234	(39-80)
235	G183	Light response; seed protein content	Dev and morph; seed biochemistry	WRKY	Constitutive photomorphogenesis; altered seed protein content	236	(307-363)
237	G2555	Light response	Dev and morph	HLH/MYC	Constitutive photomorphogenesis	238	(175-245)
239	G375	Light response	Dev and morph	Z-Dof	Upward pointing leaves	240	(75-103)
241	G1007	Morphology; other	Dev and morph	AP2	Multiple developmental alterations	242	(TBD)
243	G1010	Morphology; other	Dev and morph	ABI3/VP-1	Multiple developmental alterations	244	(33-122)
245	G1014	Morphology; other; trichome	Dev and morph	ABI3/VP-1	Multiple developmental defects; reduced trichomes	246	(90-172)
247	G1035	Morphology; other	Dev and morph	bZIP	Multiple developmental alterations	248	(39-91)
249	G1046	Morphology; other	Dev and morph	bZIP	Multiple developmental alterations	250	(79-138)
251	G1049	Morphology; other; seed protein content	Dev and morph; seed biochemistry	bZIP	Multiple developmental alterations; altered seed protein content	252	(77-132)
253	G1069	Morphology; other; seed oil content	Dev and morph; seed biochemistry	AT-hook	Multiple developmental alterations; altered seed oil content	254	(67-74)
255	G1070	Morphology; other	Dev and morph	AT-hook	Several developmental defects	256	(98-120)
257	G1076	Morphology; other	Dev and morph	AT-hook	Lethal when overexpressed	258	(82-89)
259	G1089	Morphology; other	Dev and morph	BZIP2	Developmental defects at seedling stage	260	(425-500)
261	G1093	Morphology; other	Dev and morph	RING/C3H2C3	Multiple morphological alterations	262	(105-148)

Table 4

263	G1127	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	264	(103-110, 155-162)
265	G1131	Morphology: other; seed protein content	Dev and morph; seed biochemistry	HLH/MYC	Multiple developmental alterations; altered seed protein content	266	(173-220)
267	G1145	Morphology: other; seed oil and protein	Dev and morph; seed biochemistry	bZIP	Multiple developmental alterations; reduced seed size, altered seed shape; altered seed oil and protein content	268	(227-270)
269	G1229	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	HLH/MYC	Several developmental defects; altered seed oil and protein content	270	(102-160)
271	G1246	Morphology: other; seed protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Multiple developmental alterations; altered seed protein content	272	(27-139)
273	G1255	Morphology: other; seed	Dev and morph	Z-CO-like	Reduced apical dominance; increased seed size	274	(18-56)
275	G1304	Morphology: other	Dev and morph	MYB-(R1)R2R3	Lethal when overexpressed	276	(13-118)
277	G1318	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	278	(20-123)
279	G1320	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	280	(5-108)
281	G1330	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	282	(28-134)
283	G1352	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	284	(108-129, 167-188)
285	G1354	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	286	(TBD)
287	G1360	Morphology: other	Dev and morph	NAC	Lethal when overexpressed	288	(18-174)
289	G1364	Morphology: other	Dev and morph	CAAT	Lethal when overexpressed	290	(29-120)
291	G1379	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	292	(18-85)
293	G1384	Morphology: other	Dev and morph	AP2	Abnormal inflorescence and flower development	294	(TBD)
295	G1399	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	296	(86-93)

Table 4

297	G1415	Morphology: other	Dev and morph morph; seed biochemistry	AP2	Multiple developmental alterations	298	(TBD)
299	G1417	Morphology: other; seed oil		WRKY	Reduced seedling germination and vigor; increase in 18:2, decrease in 18:3	300	(239-296)
301	G1442	Morphology: other	Dev and morph	GRF-like	Multiple developmental alterations	302	(172-223)
303	G1454	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	NAC	Multiple developmental alterations; altered seed oil and protein content	304	(9-178)
305	G1459	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	306	(10-152)
307	G1460	Morphology: other; seed protein content	Dev and morph; seed biochemistry	NAC	Multiple developmental alterations; altered seed protein content	308	(TBD)
309	G147	Morphology: other	Dev and morph	MADS	Multiple developmental defects	310	(2-57)
311	G1471	Morphology: other; seed oil	Dev and morph; seed biochemistry	Z-C2H2	Multiple developmental alterations; increased seed oil content	312	(49-70)
313	G1475	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	314	(51-73)
315	G1477	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	316	(29-48)
317	G1487	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	GATAZn	Multiple developmental alterations; altered seed oil and protein content	318	(251-276)
319	G1492	Morphology: other	Dev and morph	GARP	Multiple developmental alterations	320	(34-83)
321	G1531	Morphology: other; seed; seed protein content	Dev and morph; seed biochemistry	RING/C3HC4	Multiple developmental alterations; pale seed; altered seed protein content	322	(41-77)
323	G1540	Morphology: other	Dev and morph	HB	Reduced cell differentiation in meristem	324	(35-98)
325	G1544	Morphology: other	Dev and morph	HB	Multiple developmental alterations	326	(64-124)

Table 4

327	G156	Morphology: other; seed	Dev and morph	MADS	Multiple developmental defects; seed color alteration	328	(2-57)
329	G1584	Morphology: other	Dev and morph	HB	Multiple developmental alterations	330	(TBD)
331	G1587	Morphology: other	Dev and morph	HB	Multiple developmental alterations	332	(61-121)
333	G1588	Morphology: other	Dev and morph	HB	Multiple developmental alterations	334	(66-124)
335	G1589	Morphology: other; seed protein content	Dev and morph; seed biochemistry	HB	Multiple developmental alterations; altered seed protein content	336	(384-448)
337	G160	Morphology: other	Dev and morph	MADS	Multiple developmental defects	338	(7-62)
339	G1636	Morphology: other	Dev and morph	MYB-related	Pale green, smaller plants	340	(100-165)
341	G1642	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	342	(TBD)
343	G1747	Morphology: other; seed protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Multiple developmental alterations; altered seed protein content	344	(11-114)
345	G1749	Morphology: other	Dev and morph	AP2	Multiple developmental alterations; formation of necrotic lesions	346	(84-155)
347	G1751	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	348	(TBD)
349	G1752	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	350	(83-151)
351	G1763	Morphology: other	Dev and morph	AP2	Lethal when overexpressed	352	(140-209)
353	G1766	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	354	(10-153)
355	G1767	Morphology: other; seed oil content	Dev and morph; seed biochemistry	SCR	Multiple developmental alterations; altered seed oil content	356	(255-272)
357	G1778	Morphology: other	Dev and morph	GATA/Zn	Lethal when overexpressed	358	(94-119)
359	G1789	Morphology: other; seed protein content	Dev and morph; seed biochemistry	MYB-related	Delayed development; altered seed protein content	360	(1-50)
361	G1790	Morphology: other	Dev and morph	MYB-(R1)R2R3	Lethal when overexpressed	362	(217-316)

Table 4

363	G1791	Morphology: other	Dev and morph Dev and	AP2	Multiple developmental alterations	364	(TBD)
365	G1793	Morphology: other; seed oil	morph; seed biochemistry	AP2	Multiple developmental alterations; increased seed oil content	366	(179-255, 281-349)
367	G1795	Morphology: other; trichome	Dev and morph	AP2	Multiple developmental alterations; reduced trichomes	368	(12-80)
369	G1800	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	370	(TBD)
371	G1806	Morphology: other	Dev and morph	bZIP	Multiple developmental alterations	372	(165-225)
373	G1811	Morphology: other	Dev and morph	ABI3/VP-1	Multiple developmental alterations	374	(TBD)
375	G182	Morphology: other	Dev and morph	WRKY	Multiple developmental alterations	376	(217-276)
377	G1835	Morphology: other	Dev and morph	GATA/Zn	Small, spindly plant	378	(224-296)
379	G1836	Morphology: other	Dev and morph	CAAT	Pale green	380	(30-164)
381	G1838	Morphology: other; seed oil content	Dev and morph; seed biochemistry	AP2	Multiple developmental alterations; increased seed oil content	382	(229-305, 330-400)
383	G1843	Morphology: other	Dev and morph	MADS	Multiple developmental alterations	384	(2-57)
385	G1853	Morphology: other	Dev and morph	AKR	Lethal when overexpressed	386	(entire protein)
387	G1855	Morphology: other	Dev and morph	AKR	Slow growth	388	(entire protein)
389	G187	Morphology: other	Dev and morph	WRKY	Variety of morphological alterations	390	(172-228)
391	G1881	Morphology: other	Dev and morph	Z-CO-like	Multiple developmental alterations	392	(5-28, 56-79)
393	G1882	Morphology: other	Dev and morph	Z-Dof	Lethal when overexpressed	394	(97-125)
395	G1883	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	396	(82-124)
397	G1884	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	398	(43-71)
399	G1891	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	400	(27-69)

Table 4

401	G1896	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	402	(43-85)
403	G1898	Morphology: other	Dev and morph	Z-Dof	Lethal when overexpressed	404	(31-59)
405	G1902	Morphology: other; seed oil content	Dev and morph; seed biochemistry	Z-Dof	Multiple developmental alterations; increased seed oil content	406	(31-59)
407	G1904	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	408	(53-95)
409	G1906	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	410	(19-47)
411	G1913	Morphology: other	Dev and morph	Z-Dof	Lethal when overexpressed	412	(27-55)
413	G1914	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	414	(195-216, 245-266)
415	G1925	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	416	(6-150)
417	G1929	Morphology: other	Dev and morph	Z-CO-like	Slow growth, delayed development	418	(31-53)
419	G1930	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	420	(59-124)
421	G195	Morphology: other	Dev and morph	WRKY	Multiple developmental defects	422	(183-239)
423	G1954	Morphology: other	Dev and morph	HLH/MYC	Lethal when overexpressed	424	(187-259)
425	G1958	Morphology: other; seed protein content	Dev and morph; seed biochemistry	GARP	Reduced size and root mass in plates; altered seed protein content	426	(230-278)
427	G196	Morphology: other; seed protein content	Dev and morph; seed biochemistry	WRKY	Multiple developmental alterations; altered seed protein content	428	(223-283)
429	G1965	Morphology: other	Dev and morph	Z-Dof	Lethal when overexpressed	430	(27-55)
431	G1976	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	432	(219-323)
433	G2057	Morphology: other	Dev and morph	TEO	Multiple developmental alterations	434	(TBD)
435	G2107	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	436	(TBD)

Table 4

437	G211	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	438	(24-137)
439	G2133	Morphology: other; flowering time; seed protein content	Dev and morph; flowering time	AP2	Multiple developmental alterations; late flowering; altered seed protein content	440	(11-83)
441	G2134	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	442	(TBD)
443	G2151	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	AT-hook	Multiple developmental alterations; altered seed oil and protein content	444	(93-113, 124-144)
445	G2154	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	446	(97-119)
447	G2157	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	448	(82-102, 164-107)
449	G2181	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	450	(22-169)
451	G221	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	452	(21-125)
453	G2290	Morphology: other	Dev and morph	WRKY	Multiple developmental alterations	454	(147-205)
455	G2299	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	456	(48-115)
457	G2340	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Tissue necrosis; multiple developmental alterations; altered seed oil and protein content	458	(14-120)
459	G2346	Morphology: other	Dev and morph	SBP	Enlarged seedlings	460	(59-135)
461	G237	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	462	(11-113)
463	G2373	Morphology: other; seed protein content	Dev and morph; seed biochemistry	TH	Multiple developmental alterations; altered seed protein content	464	(290-350)
465	G2376	Morphology: other; seed oil protein	Dev and morph; seed biochemistry	TH	Seedling lethality; altered seed protein content	466	(79-178, 336-408)
467	G24	Morphology: other	Dev and morph	AP2	Reduced size and necrotic patches	468	(25-93)

Table 4

469	G2424	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	470	(107-219)
471	G2505	Morphology: other	Dev and morph	NAC	Lethal when overexpressed	472	(10-159)
473	G2512	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	474	(79-139)
475	G2513	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	476	(TBD)
477	G2519	Morphology: other	Dev and morph	HLH/MYC	Multiple developmental alterations	478	(1-65)
479	G2520	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	HLH/MYC	Multiple developmental alterations; altered seed oil and protein content	480	(135-206)
481	G2533	Morphology: other; seed protein content	Dev and morph; seed biochemistry	NAC	Multiple developmental alterations; altered seed protein content	482	(11-186)
483	G2534	Morphology: other	Dev and morph	NAC	Lethal when overexpressed	484	(10-157)
485	G2573	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Multiple developmental alterations; altered seed oil and protein content	486	(31-98)
487	G2589	Morphology: other	Dev and morph	MADS	Multiple developmental alterations	488	(2-57)
489	G2687	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	490	(51-120)
491	G27	Morphology: other	Dev and morph	AP2	Abnormal development, small	492	(37-104)
493	G2720	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Multiple developmental alterations; altered seed oil and protein content	494	(10-114)
495	G2787	Morphology: other; seed oil content	Dev and morph; seed biochemistry	AT-hook	Multiple developmental alterations; altered seed oil content	496	(172-192, 226-247, 256-276, 290-311, 245 366)
497	G2789	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	498	(53-73, 121-165)
499	G31	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	500	(TBD)

Table 4

501	G33	Morphology: other	Dev and morph	AP2	Multiple developmental defects	502	(50-117)
503	G342	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	GATA/Zn	Multiple developmental alterations; altered seed oil and protein content	504	(155-190)
505	G352	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	506	(99-119, 166-186)
507	G357	Morphology: other	Dev and morph	Z-C2H2	Developmental defect	508	(7-29)
509	G358	Morphology: other	Dev and morph	Z-C2H2	Lethal when overexpressed	510	(124-135, 188-210)
511	G360	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	512	(42-62)
513	G362	Size; Morphology: other; trichome; flowering time; seed protein content	Dev and morph; flowering time; seed biochemistry	Z-C2H2	Reduced size; increased pigmentation in seed, embryos and other organs; ectopic trichome formation; increased trichome number; late flowering; altered protein content	514	(62-82)
515	G364	Morphology: other	Dev and morph	Z-C2H2	Developmental defect	516	(54-76)
517	G365	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	518	(70-90)
519	G367	Morphology: other	Dev and morph	Z-C2H2	Lethal when overexpressed	520	(63-84)
521	G373	Morphology: other	Dev and morph	RING/C3HC4	Multiple developmental alterations	522	(129-168)
523	G396	Morphology: other; size	Dev and morph	HB	Altered leaf coloration and shape, reduced fertility; small plant	524	(159-220)
525	G431	Morphology: other	Dev and morph	HB	Developmental defect, sterile	526	(286-335)
527	G479	Morphology: other	Dev and morph	SBP	Multiple developmental alterations	528	(70-149)
529	G546	Morphology: other	Dev and morph	RING/C3H2C3	Slow growth and development; increased anthocyanin pigmentation	530	(114-155)
531	G551	Morphology: other	Dev and morph	HB	Multiple developmental alterations	532	(73-133)
533	G578	Morphology: other	Dev and morph	bZIP	Lethal when overexpressed	534	(36-96)

Table 4

535	G596	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	536	(89-96)
537	G617	Morphology: other	Dev and morph	TEO	Multiple developmental alterations	538	(64-118)
539	G620	Morphology: other; seed protein content	Dev and morph; biochemistry	CAAT	Multiple developmental alterations; altered seed protein content	540	(20-118)
541	G625	Morphology: other	Dev and morph	AP2	Lethal when overexpressed	542	(52-119)
543	G658	Morphology: other	Dev and morph	MYB-(R1)R2R3	Developmental defect	544	(2-105)
545	G716	Morphology: other	Dev and morph	ARF	Multiple developmental defects	546	(24-355)
547	G725	Morphology: other	Dev and morph	GARP	Developmental defect	548	(39-87)
549	G727	Morphology: other	Dev and morph	GARP	Multiple morphological alterations	550	(226-269)
551	G740	Morphology: other	Dev and morph	Z-CLDSH	Slow growth	552	(24-42, 232-268)
553	G770	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	554	(19-162)
555	G858	Morphology: other	Dev and morph	MADS	Multiple developmental alterations	556	(2-57)
557	G865	Morphology: other; seed protein content	Dev and morph; biochemistry	AP2	Altered morphology; increased seed protein	558	(36-103)
559	G872	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	560	(18-85)
561	G904	Morphology: other	Dev and morph	RING/C3H2C3	Multiple developmental alterations	562	(117-158)
563	G910	Morphology: other; flowering time	Dev and morph; flowering time	Z-CO-like	Multiple developmental alterations; late flowering	564	(14-37, 77-103)
565	G912	Morphology: other; size; sugar sensing; flowering time	Dev and morph; sensing; flowering time	AP2	Dark green color; small plant; reduced cotyledon expansion in glucose; late flowering	566	(51-118)
567	G920	Morphology: other	Dev and morph	WRKY	Multiple developmental alterations	568	(152-211)

Table 4

569	G939	Morphology: other; size	Dev and morph	EIL	Pale seedlings on agar; reduced size	570	(97-106)
571	G963	Morphology: other; seed	Dev and morph; biochemistry	NAC	Slowed growth rate; altered seed protein content	572	(TBD)
573	G979	Morphology: other; seed	Dev and morph	AP2	Several developmental defects; altered seed development, ripening and germination	574	(63-139, 165-233)
575	G987	Morphology: other	Dev and morph	SCR	Developmental defects	576	(428-432, 704-708)
577	G993	Morphology: other; seed	Dev and morph; biochemistry	AP2	Multiple developmental alterations; altered seed protein content	578	(69-134)
579	G681	Morphology: other; leaf	Dev and morph; biochemistry	MYB-(R1)R2R3	Multiple developmental alterations; overexpression results in an increase in M39480	580	(14-120)
581	G1482	Root	Dev and morph	Z-CO-like	Increased root growth	582	(5-63)
583	G225	Root; trichome	Dev and morph	MYB-related	Increased root hairs; glabrous, lack of trichomes	584	(39-76)
585	G226	Root; trichome; seed protein content	Dev and morph; biochemistry	MYB-related	Increased root hairs; glabrous, lack of trichomes; increased seed protein	586	(28-78)
587	G9	Root	Dev and morph	AP2	Increased root mass	588	(62-127)
589	G1040	Seed	Dev and morph	GARP	Smaller and more rounded seeds	590	(109-158)
591	G2114	Seed	Dev and morph	AP2	Increased seed size	592	(221-297, 323-393)
593	G450	Seed; size; seed protein content	Dev and morph; biochemistry	IAA	Increased seed size; reduced plant size; altered seed protein content	594	(TBD)
595	G584	Seed	Dev and morph	HLH/MYC	Large seeds	596	(401-494)
597	G668	Seed	Dev and morph	MYB-(R1)R2R3	Reduced seed color	598	(13-113)
599	G1050	Senescence	Dev and morph	bZIP	Delayed senescence	600	(372-425)
601	G1463	Senescence	Dev and morph	NAC	Premature senescence	602	(9-156)
603	G1944	Senescence; size; seed protein content	Dev and morph; biochemistry	AT-hook	Early senescence; reduced size; altered seed protein content	604	(87-100)
605	G2383	Senescence; seed protein content	Dev and morph; biochemistry	TEO	Early senescence; altered seed protein content	606	(89-149)
607	G571	Senescence; flowering time	Dev and morph; flowering time	bZIP	Delayed senescence; late flowering	608	(160-220)

Table 4

609	G636	Senescence; size	Dev and morph Dev and morph; flowering time	TH	Premature senescence; reduced size	610	(55-145, 405-498)
611	G878	Senescence; flowering time	flowering time	WRKY	Delayed senescence; late flowering	612	(250-305, 415-475)
613	G1134	Silique	Dev and morph	HLH/MYC	Siliques with altered shape	614	(198-247)
615	G1008	Size	Dev and morph	AP2	Small plant	616	(96-163)
617	G1020	Size	Dev and morph	AP2	Very small T1 plants	618	(28-95)
619	G1023	Size	Dev and morph	AP2	Reduced size	620	(128-195)
621	G1053	Size	Dev and morph	bZIP	Small plant	622	(74-120)
623	G1137	Size	Dev and morph	HLH/MYC	Small T1 plants	624	(264-314)
625	G1181	Size	Dev and morph	HS	Small T1 plants	626	(24-114)
627	G1228	Size	Dev and morph	HLH/MYC	Reduced size	628	(179-233)
629	G1277	Size	Dev and morph	AP2	Small plant	630	(18-85)
631	G1309	Size	Dev and morph	MYB-(R1)R2R3	Small plant	632	(9-114)
633	G1314	Size; sugar sensing; seed protein content	Dev and morph; sugar sensing; seed biochemistry	MYB-(R1)R2R3	Reduced size; reduced seedling vigor on high glucose; altered seed protein content	634	(14-116)
635	G1317	Size	Dev and morph	MYB-(R1)R2R3	Reduced size	636	(13-118)
637	G1323	Size; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Small T1 plants, dark green; decreased seed oil, increased seed protein	638	(15-116)
639	G1332	Size; trichome; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Reduced size; reduced trichome density; altered seed oil and protein content	640	(13-116)
641	G1334	Size	Dev and morph	CAAT	Small, dark green	642	(18-190)
643	G1381	Size	Dev and morph	AP2	Reduced size	644	(68-135)
645	G1382	Size	Dev and morph	WRKY	Small plant	646	(210-266, 385-437)
647	G1435	Size; flowering time	Dev and morph; flowering time	GARP	Increased plant size; late flowering	648	(146-194)
649	G1537	Size	Dev and morph	HB	Small T1 plants with altered development	650	(14-74)
651	G1545	Size	Dev and morph	HB	Reduced size	652	(54-117)
653	G1641	Size; seed oil and protein content	Dev and morph; seed biochemistry	MYB-related	Small plant; altered seed oil and protein content	654	(139-200)
655	G165	Size; seed protein content	Dev and morph; seed biochemistry	MADS	Reduced size; altered seed protein content	656	(7-62)

Table 4

657	G1652	Size; seed oil and protein content	Dev and morph; seed biochemistry	HLH/MYC	Reduced size; altered seed oil and protein content	658	(143-215)
659	G1655	Size	Dev and morph	HLH/MYC	Small plant	660	(134-192)
661	G1671	Size	Dev and morph	NAC	Reduced size	662	(TBD)
663	G1756	Size; seed protein content	Dev and morph; seed biochemistry	WRKY	Reduced size; altered seed protein content	664	(TBD)
665	G1757	Size; seed protein content	Dev and morph; seed biochemistry	WRKY	Small plant; altered seed protein content	666	(158-218)
667	G1782	Size	Dev and morph	CAAT	Small, spindly plant	668	(166-238)
669	G184	Size	Dev and morph	WRKY	Small plant	670	(295-352)
671	G1845	Size	Dev and morph	AP2	Small plant	672	(140-207)
673	G1879	Size; seed oil and protein content	Dev and morph; seed biochemistry	HLH/MYC	Reduced size; altered seed oil and protein content	674	(107-176)
675	G1888	Size	Dev and morph	Z-CO-like	Reduced size, dark green leaves	676	(5-50)
677	G189	Size; seed protein content	Dev and morph; seed biochemistry	WRKY	Increased leaf size; altered seed protein content	678	(240-297)
679	G1939	Size	Dev and morph	PCF	Reduced size	680	(40-102)
681	G194	Size	Dev and morph	WRKY	Small plant	682	(174-230)
683	G1943	Size	Dev and morph	HLH/MYC	Reduced size	684	(335-406)
685	G21	Size; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Reduced size; altered seed oil and protein content	686	(97-164)
687	G2132	Size; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Reduced size; altered seed oil and protein content	688	(TBD)
689	G2145	Size	Dev and morph	HLH/MYC	Reduced size	690	(166-243)
691	G23	Size	Dev and morph	AP2	Small T1 plants	692	(61-117)
693	G2313	Size	Dev and morph	MYB-related	Reduced size	694	(TBD)
695	G2344	Size	Dev and morph	CAAT	Reduced size, slow growth	696	(TBD)
697	G2430	Size	Dev and morph	GARP	Increased leaf size, faster development	698	(425-478)
699	G2517	Size	Dev and morph	WRKY	Reduced size	700	(118-234)
701	G2521	Size	Dev and morph	HLH/MYC	Reduced size	702	(145-213)
703	G258	Size; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Reduced size; altered seed oil and protein content	704	(24-124)

Table 4

705	G280	Size; seed protein content	Dev and morph; seed biochemistry	AT-hook	Reduced size; altered seed protein content	706	(97-104, 130-137-155-162, 185-192)
707	G3	Size	Dev and morph	AP2	Small plant	708	(28-95)
709	G343	Size	Dev and morph	GATA/Zn	Small plant	710	(178-214)
711	G363	Size	Dev and morph	Z-C2H2	Small plant	712	(87-108)
713	G370	Size	Dev and morph	Z-C2H2	Reduced size, shiny leaves	714	(97-117)
715	G385	Size	Dev and morph	HB	Small plant, short inflorescence stems, dark green	716	(60-123)
717	G439	Size	Dev and morph	AP2	Small plant	718	(110-177)
719	G440	Size	Dev and morph	AP2	Small plant	720	(122-189)
721	G5	Size	Dev and morph	AP2	Small plant	722	(149-216)
723	G550	Size	Dev and morph	Z-Dof	Small plant	724	(134-180)
725	G670	Size	Dev and morph	MYB-(R1)R2R3	Small plant	726	(14-122)
727	G760	Size	Dev and morph	NAC	Reduced size	728	(12-156)
729	G831	Size	Dev and morph	AKR	Reduced size	730	(470-591)
731	G864	Size	Dev and morph	AP2	Small plant	732	(119-186)
733	G884	Size	Dev and morph	WRKY	Reduced size	734	(227-285, 407-465)
735	G998	Size; seed oil and protein content	Dev and morph; seed biochemistry	RING/C3HC4	Reduced size; altered seed oil and protein content	736	(148-185)
737	G900	Size	Dev and morph	Z-CO-like	Reduced size	738	(6-28, 48-74)
739	G913	Size; flowering time	Dev and morph; flowering time	AP2	Small plant; late flowering	740	(62-128)
741	G937	Size	Dev and morph	GARP	Slightly reduced size	742	(197-246)
743	G960	Size	Dev and morph	NAC	Small plant	744	(13-156)
745	G991	Size; seed oil and protein content	Dev and morph; seed biochemistry	IAA	Slightly reduced size; altered seed oil and protein content	746	(7-14, 48-59, 82-115, 128-164)
747	G748	Stem; flowering time	Dev and morph; flowering time	Z-Dof	More vascular bundles in stem; late flowering	748	(112-140)
749	G247	Trichome; seed protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Altered trichome distribution; altered seed protein content	750	(15-116)
751	G585	Trichome	Dev and morph	HLH/MYC	Reduced trichome density	752	(436-501)
753	G634	Trichome; seed protein content	Dev and morph; seed biochemistry	TH	Increased trichome density and size; altered seed protein content	754	(62-147, 189-245)
755	G676	Trichome	Dev and morph	MYB-(R1)R2R3	Reduced trichomes	756	(17-119)

Table 4

757	G682	Trichome	Dev and morph	MYB-related	Glabrous, lack of trichomes	758	(27-63)
759	G635	Variegation	Dev and morph	TH		760	(239-323)
761	G1068	Sugar sensing;	Sugar sensing;	AT-hook	Reduced cotyledon expansion in glucose	762	(143-150)
		seed oil and protein content	seed				
763	G1225		biochemistry	HLH/MYC	Better germination on sucrose and glucose media; altered seed oil and protein content	764	(78-147)
765	G1337	Sugar sensing	Sugar sensing	Z-CO-like	Decreased germination on sucrose medium	766	(9-75)
767	G1759	Sugar sensing	Sugar sensing	MADS	Reduced germination on high glucose	768	(2-57)
		Sugar sensing;	Sugar sensing;				
769	G1804	flowering time	flowering time	bZIP	Altered sugar sensing; late flowering	770	(357-407)
771	G207	Sugar sensing	Sugar sensing	MYB-(R1)R2R3	Decreased germination on glucose medium	772	(6-106)
		Sugar sensing;	Sugar sensing;				
773	G218	seed oil content	seed	MYB-(R1)R2R3	Reduced cotyledon expansion in glucose; altered seed oil content	774	(TBD)
		Sugar sensing;	Sugar sensing;				
		seed oil and protein content	seed				
775	G241		biochemistry	MYB-(R1)R2R3	Decreased germination and growth on glucose medium; decreased seed oil, altered protein content	776	(14-114)
777	G254	Sugar sensing	Sugar sensing	MYB-related	Decreased germination and growth on glucose medium	778	(62-106)
779	G26	Sugar sensing	Sugar sensing	AP2	Decreased germination and growth on glucose medium	780	(67-134)
		Sugar sensing	Sugar sensing				
781	G263	Sugar sensing	Sugar sensing	HS	Decreased root growth on sucrose medium, root specific expression	782	(TBD)
783	G308	Sugar sensing	Sugar sensing	SCR	No germination on glucose medium	784	(270-274)
785	G38	Sugar sensing	Sugar sensing	AP2	Reduced germination on glucose medium	786	(76-143)
787	G43	Sugar sensing	Sugar sensing	AP2	Decreased germination and growth on glucose medium	788	(104-172)
789	G536	Sugar sensing	Sugar sensing	GF14	Decreased germination and growth on glucose medium	790	(226-233)
		Sugar sensing;	Sugar sensing;				
		seed oil and protein content	seed		Decreased seedling vigor on high glucose; altered seed oil and protein content	792	(210-270)
791	G567		biochemistry	bZIP			
		Sugar sensing;	Sugar sensing;				
793	G680	flowering time	flowering time	MYB-related	Reduced germination on glucose medium; late flowering	794	(24-70)
795	G667	Sugar sensing	Sugar sensing	AP2	Better seedling vigor on sucrose medium	796	(59-124)
797	G956	Sugar sensing	Sugar sensing	NAC	Reduced germination on glucose medium	798	(TBD)
799	G996	Sugar sensing	Sugar sensing	MYB-(R1)R2R3	Reduced germination on glucose medium	800	(14-114)
		Seed					
		glucosinolates, oil, protein content	Seed biochemistry		Increase in M3950; increased oil content; decreased protein content	802	(32-130)
801	G1946			HS			
		Seed oil composition	Seed biochemistry				
803	G217			MYB-related	Increase in 20:2	804	(8-67)

Table 4

805	G2192	Seed oil composition	Seed biochemistry	bZIP-NIN	Altered composition	806	(600-700)
807	G504	Seed oil composition;	Seed biochemistry	NAC	Altered seed oil composition and content; altered seed protein content	808	(TBD)
809	G622	Seed oil composition	Seed biochemistry	ABI3/NIP-1	Decreased 18:2 fatty acid	810	(TBD)
811	G778	Seed oil composition	Seed biochemistry	HLH/MYC	Increased seed 18:1 fatty acid	812	(220-267)
813	G791	Seed oil composition	Seed biochemistry	HLH/MYC	Altered seed fatty acid composition	814	(75-143)
815	G861	Seed oil composition; seed oil content	Seed biochemistry	MADS	Increase in 16:1; altered seed oil content	816	(2-57)
817	G938	Seed oil composition	Seed biochemistry	EIL	Altered seed fatty acid composition	818	(96-104)
819	G965	Seed oil composition	Seed biochemistry	HB	Increase in 18:1	820	(423-486)
821	G1143	Seed oil and protein content	Seed biochemistry	HLH/MYC	Altered seed oil and protein content	822	(33-82)
823	G1190	Seed oil content	Seed biochemistry	AKR	Increased content	824	(entire protein)
825	G1198	Seed oil and protein content	Seed biochemistry	bZIP	Altered seed oil and protein content	826	(173-223)
827	G1226	Seed oil and protein content	Seed biochemistry	HLH/MYC	Altered seed oil and protein content	828	(115-174)
829	G1451	Seed oil content	Seed biochemistry	ARF	Altered seed oil content	830	(22-357)
831	G1478	Seed oil and protein content; flowering time	Seed biochemistry; flowering time	Z-CO-like	Altered seed oil, protein content; late flowering	832	(32-76)
833	G1496	Seed oil content	Seed biochemistry	HLH/MYC	Altered seed oil content	834	(184-248)
835	G1526	Seed oil content	Seed biochemistry	SWI/SNF	Increased seed oil content	836	(493-620, 864-1006)
837	G1543	Seed oil content	Seed biochemistry	HB	Decreased seed oil	838	(135-195)
839	G162	Seed oil and protein content	Seed biochemistry	MADS	Altered seed oil content; altered seed oil and protein content	840	(2-57)

Table 4

841	G1640	Seed oil content Seed oil and protein content	Seed biochemistry	MYB-(R1)R2R3	Increased seed oil	842	(14-115)
843	G1644	Seed oil content Seed oil and protein content	Seed biochemistry	MYB-(R1)R2R3	Altered seed oil, protein content	844	(39-102)
845	G1646	Seed oil content	Seed biochemistry	CAAT	Altered seed oil content	846	(72-162)
847	G1672	Seed oil content Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil content	848	(41-194)
849	G1677	Seed oil content Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil, protein content	850	(17-181)
851	G1765	Seed oil content Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	852	(20-140)
853	G1777	Seed oil content Seed oil and protein content	Seed biochemistry	RING/C3HC4	Increased oil, decreased protein content	854	(124-247)
855	G1793	Seed oil content	Seed biochemistry	AP2	Increased seed oil content	856	(179-255, 281-349)
857	G180	Seed oil content Seed oil and protein content	Seed biochemistry	WRKY	Decreased seed oil content	858	(118-174)
859	G192	Seed oil content Seed oil and protein content	Seed biochemistry; flowering time	WRKY	Altered seed oil and protein content; late flowering	860	(128-185)
861	G1948	Seed oil content Seed oil and protein content	Seed biochemistry	AKR	Altered seed oil and protein content	862	(entire protein)
863	G2123	Seed oil content Seed oil and protein content	Seed biochemistry	GF14	Altered seed oil and protein content	864	(99-109)
865	G2138	Seed oil content	Seed biochemistry	AP2	Increased seed oil content	866	(TBD)
867	G2139	Seed oil content	Seed biochemistry	MADS	Increased seed content	868	(14-69)
869	G2343	Seed oil content Seed oil and protein content	Seed biochemistry	MYB-(R1)R2R3	Altered seed oil content	870	(14-116)
871	G265	Seed oil content Seed oil and protein content	Seed biochemistry	HS	Altered seed oil and protein content	872	(11-105)
873	G2792	Seed oil content Seed oil and protein content	Seed biochemistry	HLH/MYC	Increased seed oil content	874	(190-258)
875	G2830	Seed oil content Seed oil and protein content	Seed biochemistry	Z-C2H2	Altered seed oil and protein content	876	(245-266)
877	G286	Seed oil content	Seed biochemistry	ENBP	Altered seed oil and protein content	878	(TBD)

Table 4

879	G291	Seed oil content	Seed biochemistry	MISC	Increased seed oil content	880	(132-160)
881	G427	Seed oil and protein content	Seed biochemistry	HB	Increased oil content; decreased protein content	882	(307-370)
883	G509	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	884	(13-169)
885	G519	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	886	(11-104)
887	G561	Seed oil content	Seed biochemistry	bZIP	Altered seed oil content	888	(248-308)
889	G590	Seed oil and protein content	Seed biochemistry	HLH/MYC	Altered seed oil and protein content	890	(202-254)
891	G818	Seed oil content	Seed biochemistry	HS	Increased content	892	(70-162)
893	G849	Seed oil and protein content	Seed biochemistry	BPF-1	Increased seed oil, altered protein content	894	(324-413, 504-583)
895	G892	Seed oil and protein content	Seed biochemistry	RING/C3H2C3	Altered seed oil, protein content	896	(177-270)
897	G961	Seed oil content	Seed biochemistry	NAC	Altered seed oil content	898	(15-140)
899	G1465	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	900	(242-306)
901	G425	Seed oil content	Seed biochemistry	HB	Altered seed oil content	902	(TBD)
903	G347	Seed oil and protein content	Seed biochemistry	Z-LSDlike	Altered seed oil and protein content	904	(9-39, 50-70, 80-127)
905	G1512	Seed oil and protein content	Seed biochemistry	RING/C3HC4	Altered seed oil and protein content	906	(39-93)
907	G2069	Seed oil and protein content	Seed biochemistry	bZIP	Altered seed oil and protein content	908	(TBD)
909	G1852	Seed oil content	Seed biochemistry	AKR	Altered seed oil content	910	(1-601)
911	G1793	Seed oil content	Seed biochemistry	AP2	Altered seed oil content	912	(179-255, 281-349)
913	G761	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	914	(10-156)
915	G1056	Seed oil content	Seed biochemistry	bZIP	Altered seed oil content	916	(183-246)
917	G1447	Seed oil content	Seed biochemistry	MISC	Altered seed oil content	918	(3-54, 124-156)

Table 4

919	G323	Seed oil and protein content	Seed biochemistry	RING/C3HC4	Altered seed oil and protein content	920	(48-96)
921	G176	Seed oil content	Seed biochemistry	WRKY	Altered seed oil content	922	(117-173, 234-290)
923	G174	Seed oil and protein content	Seed biochemistry	WRKY	Altered seed oil and protein content	924	(111-166, 283-339)
925	G715	Seed oil content	Seed biochemistry	CAAT	Altered seed oil content	926	(60-132)
927	G588	Seed oil and protein content	Seed biochemistry	HLH/MYC	Altered seed oil and protein content	928	(309-376)
929	G1758	Seed oil and protein content	Seed biochemistry	WRKY	Altered seed oil and protein content	930	(109-165)
931	G2148	Seed oil content	Seed biochemistry	HLH/MYC	Altered seed oil content	932	(130-268)
933	G2379	Seed oil content	Seed biochemistry	TH	Altered seed oil content	934	(19-110, 173-232)
935	G1462	Seed oil content	Seed biochemistry	NAC	Altered seed oil content	936	(TBD)
937	G1211	Seed oil and protein content	Seed biochemistry	MISC	Altered seed oil and protein content	938	(123-179)
939	G1048	Seed oil content	Seed biochemistry	bZIP	Altered seed oil content	940	(138-190)
941	G986	Seed oil content	Seed biochemistry	WRKY	Altered seed oil content	942	(146-203)
943	G789	Seed oil content	Seed biochemistry	HLH/MYC	Altered seed oil content	944	(253-313)
945	G2085	Seed oil and protein content	Seed biochemistry	RING/C3HC4	Altered seed oil and protein content	946	(TBD)
947	G1783	Seed oil and protein content	Seed biochemistry	MYB-related	Altered seed oil and protein content	948	(81-129)
949	G2072	Seed oil and protein content	Seed biochemistry	bZIP	Altered seed oil and protein content	950	(90-149)
951	G931	Seed oil and protein content	Seed biochemistry	CAAT	Altered seed oil and protein content	952	(TBD)
953	G278	Seed oil and protein content	Seed biochemistry	AKR	Altered seed oil and protein content	954	(2-593)
955	G2421	Seed oil content	Seed biochemistry	MYB-(R1)R2R3	Altered seed oil content	956	(9-110)
957	G2032	Seed oil content	Seed biochemistry	AKR	Altered seed oil content	958	(entire protein)

Table 4

959	G1396	Seed oil and protein content	Seed biochemistry	S1FA	Altered seed oil and protein content	960	(TBD)
961	G619	Seed oil and protein content	Seed biochemistry	ARF	Altered seed oil and protein content	962	(64-406)
963	G2295	Seed oil content	Seed biochemistry	MADS	Altered seed oil content	964	(2-57)
965	G312	Seed oil content	Seed biochemistry	SCR	Altered seed oil content	966	(320-336)
967	G1444	Seed oil and protein content	Seed biochemistry	GRF-like	Altered seed oil and protein content	968	(168-193)
969	G801	Seed oil content	Seed biochemistry	PCF	Altered seed oil content	970	(32-93)
971	G1950	Seed oil content	Seed biochemistry	AKR	Altered seed oil content	972	(65-228)
973	G958	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	974	(7-156)
975	G1037	Seed oil and protein content	Seed biochemistry	GARP	Altered seed oil and protein content	976	(11-134, 200-248)
977	G2065	Seed oil content	Seed biochemistry	MADS	Altered seed oil content	978	(TBD)
979	G2137	Seed oil and protein content	Seed biochemistry	WRKY	Altered seed oil and protein content	980	(109-168)
981	G746	Seed oil content	Seed biochemistry	RING/C3HC4	Altered seed oil content	982	(139-178)
983	G2701	Seed oil and protein content	Seed biochemistry	MYB-related	Altered seed oil and protein content	984	(33-81, 129-183)
985	G1819	Seed oil content	Seed biochemistry	CAAT	Altered seed oil content	986	(46-188)
987	G1227	Seed oil and protein content	Seed biochemistry	HLH/MYC	Altered seed oil and protein content	988	(183-244)
989	G2417	Seed oil content	Seed biochemistry	GARP	Altered seed oil content	990	(235-285)
991	G2116	Seed oil content	Seed biochemistry	bZIP	Altered seed oil content	992	(150-210)
993	G647	Seed oil content	Seed biochemistry	Z-C3H	Altered seed oil content	994	(77-192)
995	G974	Seed oil and protein content	Seed biochemistry	AP2	Altered seed oil and protein content	996	(81-140)
997	G1419	Seed protein content	Seed biochemistry	AP2	Increased seed protein	998	(69-137)

Table 4

999	G1634	Seed protein content	Seed biochemistry	MYB-related	Altered seed protein content	1000	(129-180)
1001	G1637	Seed protein content	Seed biochemistry	MYB-related	Altered seed protein content	1002	(109-173)
1003	G1818	Seed protein content; flowering time	Seed biochemistry; flowering time	CAAT	Increased protein content; late flowering	1004	(36-113)
1005	G1820	Seed oil and protein content	Seed biochemistry	CAAT	Altered seed oil, protein content	1006	(70-133)
1007	G1903	Seed oil and protein content	Seed biochemistry	Z-Dof	Altered seed oil and protein content	1008	(134-180)
1009	G371	Seed oil and protein content	Seed biochemistry	RING/C3HC4	Altered seed oil and protein content	1010	(21-74)
1011	G597	Seed protein content	Seed biochemistry	AT-hook	Altered seed protein content	1012	(97-104, 137-144)
1013	G1009	Seed protein content	Seed biochemistry	AP2	Altered seed protein content	1014	(201-277, 303-371)
1015	G170	Seed protein content	Seed biochemistry	MADS	Altered seed protein content	1016	(2-57)
1017	G1768	Seed protein content	Seed biochemistry	SCR	Altered seed protein content	1018	(54-413)
1019	G185	Seed protein content	Seed biochemistry	WRKY	Altered seed protein content	1020	(113-172)
1021	G1931	Seed protein content	Seed biochemistry	WRKY	Altered seed protein content	1022	(114-170)
1023	G2543	Seed protein content	Seed biochemistry	HB	Altered seed protein content	1024	(31-91)
1025	G264	Seed protein content	Seed biochemistry	HS	Altered seed protein content	1026	(24-114)
1027	G32	Seed protein content	Seed biochemistry	AP2	Altered seed protein content	1028	(17-84)
1029	G436	Seed protein content	Seed biochemistry	HB	Altered seed protein content	1030	(22-85)
1031	G556	Seed protein content	Seed biochemistry	bZIP	Altered seed protein content	1032	(83-143)
1033	G1420	Seed protein content	Seed biochemistry	WRKY	Altered seed protein content	1034	(221-280)
1035	G1412	Seed protein content	Seed biochemistry	NAC	Altered seed protein content	1036	(17-159)

Table 4

1037	G738	Seed protein content	Seed biochemistry	Z-Dof	Altered seed protein content	1038	(351-393)
1039	G2426	Seed protein content	Seed biochemistry	MYB-(R1)R2R3	Altered seed protein content	1040	(14-114)
1041	G1524	Seed protein content	Seed biochemistry	RING/C3HC4	Altered seed protein content	1042	(49-110)
1043	G1243	Seed protein content	Seed biochemistry	SWI/SNF	Altered seed protein content	1044	(216-609)
1045	G631	Seed protein content	Seed biochemistry	bZIP	Altered seed protein content	1046	(TBD)
1047	G1909	Seed protein content	Seed biochemistry	Z-Dof	Altered seed protein content	1048	(23-51)
1049	G1663	Seed protein content	Seed biochemistry	PCF	Altered seed protein content	1050	(TBD)
1051	G1231	Seed protein content	Seed biochemistry	Z-C4HC3	Altered seed protein content	1052	(TBD)
1053	G227	Seed protein content	Seed biochemistry	MYB-(R1)R2R3	Altered seed protein content	1054	(13-112)
1055	G1842	Seed protein content	Seed biochemistry	MADS	Altered seed protein content	1056	(2-57)
1057	G1505	Seed protein content	Seed biochemistry	GATAZn	Altered seed protein content	1058	(TBD)
1059	G657	Seed protein content	Seed biochemistry	MYB-(R1)R2R3	Altered seed protein content	1060	(TBD)
1061	G1959	Seed protein content	Seed biochemistry	GARP	Altered seed protein content	1062	(46-97)
1063	G2180	Seed protein content	Seed biochemistry	NAC	Altered seed protein content	1064	(7-156)
1065	G1817	Seed protein content	Seed biochemistry	PMR	Altered seed protein content	1066	(47-331)
1067	G1649	Seed protein content	Seed biochemistry	HLH/MYC	Altered seed protein content	1068	(225-295)
1069	G2131	Seed protein content	Seed biochemistry	AP2	Altered seed protein content	1070	(50-186, 112-183)
1071	G215	Seed protein content	Seed biochemistry	MYB-related	Altered seed protein content	1072	(TBD)
1073	G1508	Seed protein content	Seed biochemistry	GATAZn	Altered seed protein content	1074	(38-63)
1075	G2110	Seed protein content	Seed biochemistry	WRKY	Altered seed protein content	1076	(239-298)

Table 4

1077	G2442	Seed protein content	Seed biochemistry	RING/C3HC4	Altered seed protein content	1078	(220-246)
1079	G1051	Flowering time	Flowering time	bZIP	Late flowering	1080	(189-250)
1081	G1052	Flowering time	Flowering time	bZIP	Late flowering	1082	(201-261)
1083	G1079	Flowering time;	Flowering time;	BZIP2	Late flowering; altered seed protein content	1084	(1-50)
1085	G1335	Flowering time	Flowering time	Z-CLDSH	Late flowering; slow growth	1086	(24-43, 131-144, 185-203)
1087	G157	Flowering time	Flowering time	MADS	Altered flowering; significant overexpression delays	1088	(2-57)
1089	G1895	Flowering time	Flowering time	Z-Dof	Late flowering	1090	(55-110)
1091	G1900	Flowering time	Flowering time	Z-Dof	Late flowering	1092	(54-106)
1093	G2007	Seed protein content	Seed biochemistry	MYB-(R1)R2R3	Late flowering; altered seed protein content	1094	(TBD)
1095	G214	Flowering time	Flowering time	MYB-related	Late flowering	1096	(22-71)
1097	G2155	Flowering time	Flowering time	AT-hook	Late flowering	1098	(18-38)
1099	G234	Flowering time	Flowering time	MYB-(R1)R2R3	Late flowering; small plant	1100	(14-115)
1101	G361	Flowering time	Flowering time	Z-C2H2	Late flowering	1102	(43-63)
1103	G562	Flowering time	Flowering time	bZIP	Late flowering	1104	(253-315)
1105	G591	Flowering time	Flowering time	HLH/MYC	Late flowering	1106	(143-240)
1107	G8	Flowering time	Flowering time	AP2	Late flowering	1108	(151-217, 243-296)
1109	G859	Seed protein content	Seed biochemistry	MADS	Late flowering; altered seed protein content	1110	(TBD)
1111	G878	Flowering time	Flowering time	WRKY	Late flowering	1112	(250-305, 415-475)
1113	G971	Flowering time	Flowering time	AP2	Late flowering	1114	(120-186)
1115	G975	Flowering time; morphology: other	Flowering time; dev and morph	AP2	Late flowering; glossy leaves	1116	(4-71)
1117	G994	Flowering time	Flowering time	MYB-(R1)R2R3	Late flowering; small	1118	(14-123)
1119	G2347	Flowering time	Flowering time	SBP	Late flowering; small	1120	(60-136)
1121	G2010	Flowering time	Flowering time	SBP	Late flowering	1122	(53-127)

Table 5

SEQ ID NO	GID	Test Sequence ID	Smallest Sum Probability	Test Sequence Species	Test Sequence GenBank Annotation
859	G192	AW596933	7.70E-40	[Glycine max]	sj84f07.y1 Gm-c1034 Glycine max cDNA clone GENO
859	G192	AV423663	2.40E-39	[Lotus japonicus]	AV423663 Lotus japonicus young plants (two-
859	G192	BI422074	4.50E-34	[Lycopersicon esculentum]	EST532740 tomato callus, TAMU Lycop
859	G192	AW447931	1.40E-27	[Triticum aestivum]	BRY 1082 BRY Triticum aestivum cDNA clone
859	G192	BE998060	2.60E-24	[Medicago truncatula]	EST429783 GVSN Medicago truncatula cDNA
859	G192	AC018727	1.70E-23	[Oryza sativa]	chromosome 10 clone OSJNBa0056G17, *** SEQUENC
859	G192	BG600477	1.00E-20	[Solanum tuberosum]	EST505372 cSTS Solanum tuberosum cDNA clo
859	G192	BG356878	2.80E-16	[Sorghum bicolor]	OV2 11 B04.g1 A002 Ovary 2 (OV2) Sorghum bi
859	G192	gi12039364	1.10E-31	[Oryza sativa]	putative DNA-binding protein.
859	G192	gi4894963	3.30E-14	[Avena sativa]	DNA-binding protein WRKY3.
859	G192	gi1432056	5.80E-14	[Petroselinum crispum]	WRKY3.
859	G192	gi4760596	2.60E-13	[Nicotiana tabacum]	DNA-binding protein NtWRKY3.
859	G192	gi11993901	1.40E-12	[Dactylis glomerata]	somatic embryogenesis related protein.
859	G192	gi927025	7.60E-09	[Cucumis sativus]	SPF1-like DNA-binding protein.
859	G192	gi13620227	8.40E-09	[Lycopersicon esculentum]	hypothetical protein.
859	G192	gi3420906	2.80E-08	[Pimpinella brachycarpa]	zinc finger protein; WRKY1.
859	G192	gi1159877	4.70E-08	[Avena fatua]	DNA-binding protein.
859	G192	gi484261	1.60E-07	[Ipomoea batatas]	SPF1 protein.
801	G1946	LPHSF8	1.10E-119	[Lycopersicon peruvianum]	L.peruvianum Lp-hsf8 mRNA for heat
801	G1946	AC087771	4.10E-112	[Medicago truncatula]	clone 8D15, *** SEQUENCING IN PROGRESS
801	G1946	LEHSF8	5.90E-103	[Lycopersicon esculentum]	L.esculentum Le-hsf8 gene for heat
801	G1946	AW569138	3.10E-75	[Glycine max]	si63g09.y1 Gm-r1030 Glycine max cDNA clone GENO
801	G1946	BG890899	1.30E-70	[Solanum tuberosum]	EST516750 cSTD Solanum tuberosum cDNA clo
801	G1946	AC027658	4.60E-53	[Oryza sativa]	subsp. japonica BAC nbxb0006113, chromosome 10
801	G1946	AV833112	4.90E-52	[Hordeum vulgare subsp. vulgare]	AV833112 K. Sato unpublished
801	G1946	gi19492	2.80E-121	[Lycopersicon peruvianum]	heat shock transcription factor 8
801	G1946	gi19260	5.10E-106	[Lycopersicon esculentum]	heat stress transcription factor
801	G1946	gi662924	2.00E-47	[Glycine max]	heat shock transcription factor 21.
801	G1946	gi5821138	9.70E-46	[Nicotiana tabacum]	heat shock factor.
801	G1946	gi11761077	2.90E-40	[Oryza sativa]	putative heat shock factor protein 1 (HSF 1)
801	G1946	gi886742	3.20E-40	[Zea mays]	heat shock factor.
801	G1946	gi7158882	2.70E-38	[Medicago sativa]	heat shock transcription factor.
801	G1946	gi3550588	1.90E-30	[Pisum sativum]	heat shock transcription factor (HSFA).

Table 5

801	G1946	gi100546	0.46	[Avena sativa]	avenin precursor - oat.
801	G1946	gi114190783	1	[Apium graveolens]	putative phloem transcription factor M1.
239	G375	AW696439	3.40E-33	[Medicago truncatula]	NF106B07ST1F1060 Developing stem Medica
239	G375	BG595870	1.90E-31	[Solanum tuberosum]	EST494548 cSTS Solanum tuberosum cDNA clo
239	G375	AI899263	3.70E-31	[Lycopersicon esculentum]	EST268706 tomato ovary, TAMU Lycope
239	G375	NTBBF3	4.00E-31	[Nicotiana tabacum]	N.tabacum mRNA for zinc finger protein, B
239	G375	BG405482	2.70E-30	[Glycine max]	sac44a11.y1 Gm-c1062 Glycine max cDNA clone GEN
239	G375	AB028130	3.30E-30	[Oryza sativa]	mRNA for Dof zinc finger protein, complete cds
239	G375	AB026297	7.30E-28	[Pisum sativum]	mRNA for elicitor-responsive Dof protein ERDP
239	G375	HVBPF	1.10E-27	[Hordeum vulgare]	mRNA for DNA binding protein BPBF.
239	G375	BG263089	1.70E-27	[Triticum aestivum]	WHE2337_A02_A03ZS Wheat pre-anthesis spik
239	G375	ZMU82230	4.20E-27	[Zea mays]	endosperm-specific prolamin box binding factor (PB
239	G375	gi4996640	1.90E-37	[Oryza sativa]	Dof zinc finger protein.
239	G375	gi3777436	8.10E-35	[Hordeum vulgare]	DNA binding protein.
239	G375	gi2393775	1.10E-33	[Zea mays]	prolamin box binding factor.
239	G375	gi1360088	2.00E-33	[Nicotiana tabacum]	Zn finger protein.
239	G375	gi3790264	4.30E-32	[Triticum aestivum]	PBF protein.
239	G375	gi6092016	1.30E-29	[Pisum sativum]	elicitor-responsive Dof protein ERDP.
239	G375	gi7688355	5.60E-29	[Solanum tuberosum]	Dof zinc finger protein.
239	G375	gi1669341	4.60E-20	[Cucurbita maxima]	AOBP (ascorbate oxidase promoter-binding
239	G375	gi3929325	5.50E-18	[Dendrobium gress Madama Thong-In]	putative DNA-binding prot
239	G375	gi19547	5.50E-06	[Medicago sativa subsp. falcata]	environmental stress and a
273	G1255	AC087181	1.60E-46	[Oryza sativa]	chromosome 3 clone OSJNBa0018H01, *** SEQUENCI
273	G1255	BG239774	4.50E-33	[Glycine max]	sab74c03.y1 Gm-c1032 Glycine max cDNA clone GEN
273	G1255	BG321336	1.70E-32	[Descurainia sophia]	Ds01_06h10_A Ds01_AAFECORC_cold_stress
273	G1255	AI772841	2.90E-30	[Lycopersicon esculentum]	EST253941 tomato resistant, Cornell
273	G1255	BF480245	4.60E-29	[Mesembryanthemum crystallinum]	L0-2152T3 Ice plant Lambda Un
273	G1255	AW688119	2.10E-28	[Medicago truncatula]	NF002E07ST1F1000 Developing stem Medica
273	G1255	BF266327	1.80E-26	[Hordeum vulgare]	HV_CEA0014N02f Hordeum vulgare seedling gre
273	G1255	AW671538	5.80E-25	[Sorghum bicolor]	LG1_348_B08.b1_A002 Light Grown 1 (LG1) Sor
273	G1255	BI072021	5.30E-20	[Populus tremula x Populus tremuloides]	C067P76U Populus stra
273	G1255	BG273908	4.90E-19	[Vitis vinifera]	EST 110 Green Grape berries Lambda Zap II Li
273	G1255	gi13702811	3.70E-52	[Oryza sativa]	putative zinc finger protein.
273	G1255	gi11037311	4.00E-21	[Brassica nigra]	constans-like protein.
273	G1255	gi2303683	1.10E-19	[Brassica napus]	unnamed protein product.
273	G1255	gi4091804	2.30E-18	[Malus x domestica]	CONSTANS-like protein 1.

Table 5

273	G1255	gi3341723	4.30E-18	[Raphanus sativus]	CONSTANS-like 1 protein.
273	G1255	gi10946337	5.20E-17	[Ipomoea nil]	CONSTANS-like protein.
273	G1255	gi4557093	3.30E-15	[Pinus radiata]	zinc finger protein.
273	G1255	gi8132543	0.97	[Chloroplast Zamia furfuracea]	cytochrome b559 alpha subunit
273	G1255	gi11795	0.99	[Nicotiana tabacum]	put. psbE protein (aa 1-83).
273	G1255	gi65646	0.99	[Chloroplast Nicotiana tabacum]	cytochrome b559 component p
557	G865	BE419451	3.70E-32	[Triticum aestivum]	WWS012.C2R000101 ITEC WWS Wheat Scutellum
557	G865	AW560968	1.10E-28	[Medicago truncatula]	EST316016 DSIR Medicago truncatula cDNA
557	G865	AW782252	1.20E-26	[Glycine max]	sm03d11.y1 Gm-c1027 Glycine max cDNA clone GENO
557	G865	BI421895	3.60E-25	[Lycopersicon esculentum]	EST532561 tomato callus, TAMU Lycopersicon
557	G865	BE642320	1.60E-24	[Ceratopteris richardii]	Gr2_5_L17_SP6 Ceratopteris Spore Li
557	G865	BE494041	1.60E-24	[Secale cereale]	WHE1277_B09_D17ZS Secale cereale anther cDNA
557	G865	D39914	2.60E-24	[Oryza sativa]	RICS1576A Rice shoot Oryza sativa cDNA, mRNA s
557	G865	AV428124	9.00E-23	[Lotus japonicus]	AV428124 Lotus japonicus young plants (two-)
557	G865	TOBBY4D	1.80E-21	[Nicotiana tabacum]	Tobacco mRNA for EREBP-2, complete cds.
557	G865	gi1208495	2.40E-23	[Nicotiana tabacum]	ERF1.
557	G865	gi8809571	5.10E-23	[Nicotiana sylvestris]	ethylene-responsive element binding
557	G865	gi3342211	1.40E-22	[Lycopersicon esculentum]	Pti4.
557	G865	gi7528276	1.70E-22	[Mesembryanthemum crystallinum]	AP2-related transcription f
557	G865	gi15217291	7.80E-22	[Oryza sativa]	Putative AP2 domain containing protein.
557	G865	gi3264767	2.70E-21	[Prunus armeniaca]	AP2 domain containing protein.
557	G865	gi8980313	2.10E-20	[Catharanthus roseus]	AP2-domain DNA-binding protein.
557	G865	gi8571476	9.30E-20	[Atriplex hortensis]	apetala2 domain-containing protein.
557	G865	gi1688233	1.40E-19	[Solanum tuberosum]	DNA binding protein homolog.
557	G865	gi6478845	1.80E-19	[Mairicaria chamomilla]	ethylene-responsive element binding
23	G2509	BH577856	2.50E-29	[Brassica oleracea]	BOHOJ67TR BOHO Brassica oleracea genomic
23	G2509	BM269574	5.90E-28	[Glycine max]	sak01e08.y1 Gm-c1074 Glycine max cDNA clone SOY
23	G2509	BE419451	2.20E-27	[Triticum aestivum]	WWS012.C2R000101 ITEC WWS Wheat Scutellum
23	G2509	AI483636	7.80E-27	[Lycopersicon esculentum]	EST249507 tomato ovary, TAMU Lycopersicon
23	G2509	AW560968	8.90E-27	[Medicago truncatula]	EST316016 DSIR Medicago truncatula cDNA
23	G2509	BE642320	4.30E-26	[Ceratopteris richardii]	Gr2_5_L17_SP6 Ceratopteris Spore Li
23	G2509	AP003286	1.00E-25	[Oryza sativa]	chromosome 1 clone P0677H08, *** SEQUENCING IN
23	G2509	BE494041	3.20E-25	[Secale cereale]	WHE1277_B09_D17ZS Secale cereale anther cDNA
23	G2509	BE602106	1.10E-24	[Hordeum vulgare]	HVSMEH0102106f Hordeum vulgare 5-45 DAP spi
23	G2509	AV428124	1.00E-23	[Lotus japonicus]	AV428124 Lotus japonicus young plants (two-)
23	G2509	gi3264767	4.00E-27	[Prunus armeniaca]	AP2 domain containing protein.

Table 5

23	G2509	gi12003376	1.40E-23	[Nicotiana tabacum]	Avr/Cf-9 rapidly elicited protein 1.
23	G2509	gi14140141	2.30E-23	[Oryza sativa]	putative AP2-related transcription factor.
23	G2509	gi1688233	5.40E-23	[Solanum tuberosum]	DNA binding protein homolog.
23	G2509	gi4099921	2.60E-22	[Stylosanthes hamata]	EREBP-3 homolog.
23	G2509	gi8809571	7.80E-22	[Nicotiana sylvestris]	ethylene-responsive element binding
23	G2509	gi3342211	1.00E-21	[Lycopersicon esculentum]	Pt4.
23	G2509	gi7528276	2.70E-21	[Mesembryanthemum crystallinum]	AP2-related transcription f
23	G2509	gi17385636	1.90E-20	[Matricaria chamomilla]	ethylene-responsive element binding
23	G2509	gi18496063	3.30E-20	[Fagus sylvatica]	ethylene responsive element binding prote
1119	G2347	BI931517	5.30E-31	[Lycopersicon esculentum]	EST551406 tomato flower, 8 mm to pr
1119	G2347	BE058432	4.20E-29	[Glycine max]	sn16a06.y1 Gm-c1016 Glycine max cDNA clone GENO
1119	G2347	AMSPB1	1.80E-28	[Antirrhinum majus]	A.majus mRNA for squamosa-promoter bindin
1119	G2347	BG525285	5.70E-28	[Stevia rebaudiana]	48-3 Stevia field grown leaf cDNA Stevia
1119	G2347	L38193	4.60E-27	[Brassica rapa]	BNAF1025E Mustard flower buds Brassica rapa c
1119	G2347	BG455868	6.40E-27	[Medicago truncatula]	NF068F05PL1F1045 Phosphate starved leaf
1119	G2347	BG097153	1.70E-24	[Solanum tuberosum]	EST461672 potato leaves and petioles Sola
1119	G2347	BF482644	1.60E-23	[Triticum aestivum]	WHE2301-2304_A21_A21ZS Wheat pre-anthesis
1119	G2347	AW747167	2.30E-23	[Sorghum bicolor]	WS1_66_F11.b1_A002 Water-stressed 1 (WS1) S
1119	G2347	BG442540	2.50E-23	[Gossypium arboreum]	GA_Ea0017G06f Gossypium arboreum 7-10 d
1119	G2347	gi1183864	1.50E-31	[Antirrhinum majus]	squamosa-promoter binding protein 2.
1119	G2347	gi5931786	3.40E-25	[Zea mays]	SBP-domain protein 5.
1119	G2347	gi8468036	1.40E-21	[Oryza sativa]	Similar to Arabidopsis thaliana chromosome 2
1119	G2347	gi9087308	6.60E-09	[Mitochondrion Beta vulgaris var. altissima]	orf102a.
1119	G2347	gi7209500	0.83	[Brassica rapa]	S-locus pollen protein.
43	G988	CRU303349	3.10E-208	[Capsella rubella]	ORF1, ORF2, ORF3, ORF4, ORF5 and ORF6 (pa
43	G988	A84072	4.50E-86	[Lycopersicon esculentum]	Sequence 1 from Patent WO9846759.
43	G988	A84080	3.30E-85	[Solanum tuberosum]	Sequence 9 from Patent WO9846759.
43	G988	AP003944	1.30E-57	[Oryza sativa]	chromosome 6 clone OJ1126_F05, *** SEQUENCING
43	G988	AX081276	2.80E-43	[Brassica napus]	Sequence 1 from Patent WO01093556.
43	G988	ZMA242530	1.50E-40	[Zea mays]	partial d8 gene for gibberellin response modulator
43	G988	AX005804	2.50E-37	[Triticum aestivum]	Sequence 13 from Patent WO9909174.
43	G988	AB048713	9.10E-33	[Pisum sativum]	PsSCR mRNA for SCARECROW, complete cds.
43	G988	AW774515	2.00E-29	[Medicago truncatula]	EST333666 KV3 Medicago truncatula cDNA
43	G988	BE822458	1.20E-27	[Glycine max]	GM700017A20H12 Gm-r1070 Glycine max cDNA clone
43	G988	gi13620166	8.00E-211	[Capsella rubella]	hypothetical protein.
43	G988	gi4160441	1.40E-87	[Lycopersicon esculentum]	lateral suppressor protein.

Table 5

43	G988	gi10178637	2.20E-48	[Zea mays]	SCARECROW.
43	G988	gi6970472	1.20E-47	[Oryza sativa]	OsGAI.
43	G988	gi5640157	2.80E-45	[Triticum aestivum]	gibberellin response modulator.
43	G988	gi13170126	7.10E-45	[Brassica napus]	unnamed protein product.
43	G988	gi13365610	1.10E-40	[Pisum sativum]	SCARECROW.
43	G988	gi14318115	1.10E-14	[Zea mays subsp. mays]	gibberellin response modulator.
43	G988	gi14318165	7.30E-14	[Tripsacum dactyloides]	gibberellin response modulator.
43	G988	gi347457	2.40E-05	[Glycine max]	hydroxyproline-rich glycoprotein.
459	G2346	AMA011622	3.10E-35	[Antirrhinum majus]	mRNA for squamosa promoter binding
459	G2346	AW691786	1.80E-26	[Medicago truncatula]	NF044B06ST1F1000 Developing stem Medica
459	G2346	AQ273505	7.00E-25	[Oryza sativa]	nbxb0030003f CUGI Rice BAC Library Oryza sativ
459	G2346	AW932595	7.90E-24	[Lycopersicon esculentum]	EST358438 tomato fruit mature green
459	G2346	BG593787	9.50E-24	[Solanum tuberosum]	EST492465 cSTS Solanum tuberosum cDNA clo
459	G2346	BG442540	1.00E-23	[Gossypium arboreum]	GA Ea0017G06f Gossypium arboreum 7-10 d
459	G2346	AZ919034	1.90E-23	[Zea mays]	1006013G02.x3 1006 - RescueMu Grid G Zea mays geno
459	G2346	BE596165	2.70E-23	[Sorghum bicolor]	PI1 50 D04.b1 A002 Pathogen induced 1 (PI1)
459	G2346	AI443033	2.30E-22	[Glycine max]	sa31a08.y1 Gm-c1004 Glycine max cDNA clone GENO
459	G2346	BF482644	4.30E-22	[Triticum aestivum]	WHE2301-2304 A21 A21ZS Wheat pre-anthesis
459	G2346	gi5931643	6.20E-45	[Antirrhinum majus]	squamosa promoter binding protein-homol
459	G2346	gi5931786	4.20E-26	[Zea mays]	SBP-domain protein 5.
459	G2346	gi8468036	3.30E-14	[Oryza sativa]	Similar to Arabidopsis thaliana chromosome 2
459	G2346	gi9087308	8.30E-08	[Mitochondrion Beta vulgaris var. altissima]	orf102a.
285	G1354	BG128374	2.90E-58	[Lycopersicon esculentum]	EST474020 tomato shoot/meristem Lyc
285	G1354	BE202831	1.90E-56	[Medicago truncatula]	EST402853 KV1 Medicago truncatula cDNA
285	G1354	AI161918	6.60E-55	[Populus tremula x Populus tremuloides]	A009P50U Hybrid aspen
285	G1354	AB028186	1.20E-53	[Oryza sativa]	mRNA for OsNAC7 protein, complete cds.
285	G1354	BE060921	8.00E-50	[Hordeum vulgare]	HVSMeg0013N15f Hordeum vulgare pre-anthesis
285	G1354	AF402603	1.50E-42	[Phaseolus vulgaris]	NAC domain protein NAC2 mRNA, complete c
285	G1354	BE357920	1.60E-42	[Sorghum bicolor]	DG1 23 F03.b1 A002 Dark Grown 1 (DG1) Sorgh
285	G1354	PHRNANAM	3.60E-42	[Petunia x hybrida]	P.hybrida mRNA encoding NAM protein.
285	G1354	AW185617	5.30E-40	[Glycine max]	se80b05.y1 Gm-c1023 Glycine max cDNA clone GENO
285	G1354	gi6006373	4.50E-63	[Oryza sativa]	Similar to NAM like protein (AC005310).
285	G1354	gi15148914	2.30E-44	[Phaseolus vulgaris]	NAC domain protein NAC2.
285	G1354	gi14485513	3.50E-44	[Solanum tuberosum]	putative NAC domain protein.
285	G1354	gi1279640	5.90E-44	[Petunia x hybrida]	NAM.
285	G1354	gi6175246	5.20E-41	[Lycopersicon esculentum]	jasmonic acid 2.

Table 5

285	G1354	gi4218535	5.10E-39	[Triticum sp.]	GRAB1 protein.
285	G1354	gi6732158	5.10E-39	[Triticum monococcum]	unnamed protein product.
285	G1354	gi7716952	3.30E-35	[Medicago truncatula]	NAC1.
285	G1354	gi4996349	2.50E-26	[Nicotiana tabacum]	NAC-domain protein.
285	G1354	gi2982275	3.10E-14	[Picea mariana]	ATAF1-like protein.
119	G1063	BH700922	4.50E-90	[Brassica oleracea]	BOMMZ07TR BO 2 3 KB Brassica oleracea gen
119	G1063	BE451174	2.40E-41	[Lycopersicon esculentum]	EST402062 tomato root, plants pre-a
119	G1063	AW832545	2.00E-40	[Glycine max]	sm12e10.y1 Gm-c1027 Glycine max cDNA clone GENO
119	G1063	AP004693	5.90E-37	[Oryza sativa]	chromosome 8 clone P0461F06, *** SEQUENCING IN
119	G1063	AP004462	4.40E-32	[Oryza sativa (japonica cultivar-group)]	() chromosome 8 clo
119	G1063	AT002234	8.90E-32	[Brassica rapa subsp. pekinensis]	AT002234 Flower bud cDNA Br
119	G1063	BF263465	5.40E-25	[Hordeum vulgare]	HV CEa0006N02f Hordeum vulgare seedling gre
119	G1063	BG557011	4.20E-22	[Sorghum bicolor]	EM1 41 E02.g1 A002 Embryo 1 (EM1) Sorghum b
119	G1063	BG842856	3.10E-21	[Zea mays]	MEST40-H05.T3 ISUM4-TN Zea mays cDNA clone MEST40-
119	G1063	BG559930	1.40E-18	[Sorghum propinquum]	RHIZ2 75 D09.g1 A003 Rhizome2 (RHIZ2) So
119	G1063	gi15528743	4.20E-26	[Oryza sativa]	contains EST C74560(E31855)-unknown protein.
119	G1063	gi6166283	8.10E-10	[Pinus taeda]	helix-loop-helix protein 1A.
119	G1063	gi11045087	8.80E-09	[Brassica napus]	putative protein.
119	G1063	gi10998404	7.10E-08	[Petunia x hybrida]	anthocyanin 1.
119	G1063	gi99441	2.60E-07	[Volvox carteri]	sulfated surface glycoprotein 185 - Volvox
119	G1063	gi1142621	5.00E-07	[Phaseolus vulgaris]	phaseolin G-box binding protein PG2.
119	G1063	gi166428	8.10E-07	[Antirrhinum majus]	DEL.
119	G1063	gi1247386	9.50E-07	[Nicotiana glauca]	PRP2.
119	G1063	gi82091	1.00E-06	[Lycopersicon esculentum]	hydroxyproline-rich glycoprotein
119	G1063	gi1486263	1.40E-06	[Catharanthus roseus]	extensin.
129	G2143	BH650724	3.00E-88	[Brassica oleracea]	BOMIW43TR BO 2 3 KB Brassica oleracea gen
129	G2143	AW832545	1.50E-40	[Glycine max]	sm12e10.y1 Gm-c1027 Glycine max cDNA clone GENO
129	G2143	BE451174	3.50E-40	[Lycopersicon esculentum]	EST402062 tomato root, plants pre-a
129	G2143	AP004693	4.00E-38	[Oryza sativa]	chromosome 8 clone P0461F06, *** SEQUENCING IN
129	G2143	AP004584	6.30E-33	[Oryza sativa (japonica cultivar-group)]	() chromosome 8 clo
129	G2143	AT002234	3.00E-31	[Brassica rapa subsp. pekinensis]	AT002234 Flower bud cDNA Br
129	G2143	BF263465	2.90E-26	[Hordeum vulgare]	HV CEa0006N02f Hordeum vulgare seedling gre
129	G2143	BG557011	2.60E-22	[Sorghum bicolor]	EM1 41 E02.g1 A002 Embryo 1 (EM1) Sorghum b
129	G2143	BG842856	3.50E-20	[Zea mays]	MEST40-H05.T3 ISUM4-TN Zea mays cDNA clone MEST40-
129	G2143	BG559930	6.10E-18	[Sorghum propinquum]	RHIZ2 75 D09.g1 A003 Rhizome2 (RHIZ2) So
129	G2143	gi15528743	5.50E-26	[Oryza sativa]	contains EST C74560(E31855)-unknown protein.

Table 5

129	G2143	gi1086538	7.60E-09	[Oryza rufipogon]	transcriptional activator Rb homolog.
129	G2143	gi6166283	1.10E-08	[Pinus taeda]	helix-loop-helix protein 1A.
129	G2143	gi1142621	4.60E-07	[Phaseolus vulgaris]	phascolin G-box binding protein PG2.
129	G2143	gi3399777	5.20E-07	[Glycine max]	symbiotic ammonium transporter; nodulin.
129	G2143	gi5923912	6.10E-07	[Tulipa gesneriana]	bHLH transcription factor GBOF-1.
129	G2143	gi10998404	9.20E-07	[Petunia x hybrida]	anthocyanin 1.
129	G2143	gi4321762	5.20E-06	[Zea mays]	transcription factor MYC7E.
129	G2143	gi166428	6.00E-06	[Antirrhinum majus]	DEL.
129	G2143	gi527665	7.40E-06	[Sorghum bicolor]	myc-like regulatory R gene product.
133	G2557	BH511840	6.70E-62	[Brassica oleracea]	BOGRJ19TR BOGR Brassica oleracea genomic
133	G2557	BE347811	3.70E-46	[Glycine max]	sp05h10.y1 Gm-c1041 Glycine max cDNA clone GENO
133	G2557	AP003141	2.40E-33	[Oryza sativa]	genomic DNA, chromosome 1, PAC clone:P0002B05,
133	G2557	BF263465	3.00E-31	[Hordeum vulgare]	HV_CEA0006N02f Hordeum vulgare seedling gre
133	G2557	AT002234	6.60E-27	[Brassica rapa subsp. pekinensis]	AT002234 Flower bud cDNA Br
133	G2557	BG557011	6.40E-26	[Sorghum bicolor]	EM1_41_E02.g1_A002 Embryo 1 (EM1) Sorghum b
133	G2557	AP004462	7.90E-26	[Oryza sativa (japonica cultivar-group)]	() chromosome 8 clob
133	G2557	BE451174	3.90E-25	[Lycopersicon esculentum]	EST402062 tomato root, plants pre-a
133	G2557	BG842856	5.60E-22	[Zea mays]	MEST40-H05.T3 ISUM4-TN Zea mays cDNA clone MEST40-
133	G2557	BG559930	7.00E-14	[Sorghum propinquum]	RHIZ2_75_D09.g1_A003 Rhizome2 (RHIZ2) So
133	G2557	gi15289790	2.40E-36	[Oryza sativa]	contains EST_C74560(E31855)-unknown protein.
133	G2557	gi3399777	2.60E-06	[Glycine max]	symbiotic ammonium transporter; nodulin.
133	G2557	gi4206118	1.10E-05	[Mesembryanthemum crystallinum]	transporter homolog.
133	G2557	gi6166283	1.30E-05	[Pinus taeda]	helix-loop-helix protein 1A.
133	G2557	gi527655	3.70E-05	[Pennisetum glaucum]	myc-like regulatory R gene product.
133	G2557	gi5923912	3.70E-05	[Tulipa gesneriana]	bHLH transcription factor GBOF-1.
133	G2557	gi527661	7.80E-05	[Phyllostachys acuta]	myc-like regulatory R gene product.
133	G2557	gi527665	9.50E-05	[Sorghum bicolor]	myc-like regulatory R gene product.
133	G2557	gi1086538	0.0001	[Oryza rufipogon]	transcriptional activator Rb homolog.
133	G2557	gi5669656	0.00013	[Lycopersicon esculentum]	ER33 protein.
697	G2430	BF632520	1.90E-14	[Medicago truncatula]	NF039A08DT1F1054 Drought Medicago trunc
697	G2430	AW396912	1.20E-13	[Glycine max]	sg64g09.y1 Gm-c1007 Glycine max cDNA clone GENO
697	G2430	D41804	4.50E-13	[Oryza sativa]	RICS4626A Rice shoot Oryza sativa cDNA, mRNA s
697	G2430	BE214029	2.60E-10	[Hordeum vulgare]	HV_CEB0001P06f Hordeum vulgare seedling gre
697	G2430	AW564570	2.70E-10	[Sorghum bicolor]	LG1_296_E01.b1_A002 Light Grown 1 (LG1) Sor
697	G2430	BG129795	5.40E-10	[Lycopersicon esculentum]	EST475441 tomato shoot/meristem Lyc
697	G2430	AB060130	5.40E-09	[Zea mays]	ZmRR8 mRNA for response regulator 8, complete cds.

Table 5

697	G2430	BF587105	2.50E-05	[Sorghum propinquum]	FM1_32_C05.b1_A003 Floral-Induced Merist
697	G2430	A1163121	0.3	[Populus tremula x Populus tremuloides]	A033P70U Hybrid aspen
697	G2430	BG595628	0.46	[Solanum tuberosum]	EST494306 cSTS Solanum tuberosum cDNA clo
697	G2430	gi13661174	5.40E-18	[Zea mays]	response regulator 8.
697	G2430	gi15289981	0.028	[Oryza sativa]	hypothetical protein.
697	G2430	gi6942190	0.12	[Mesembryanthemum crystallinum]	CDPK substrate protein 1; C
697	G2430	gi4519671	0.2	[Nicotiana tabacum]	transfactor.
831	G1478	BF275913	1.50E-20	[Gossypium arboreum]	GA_Eb0025C07f Gossypium arboreum 7-10 d
831	G1478	BG157399	6.50E-19	[Glycine max]	sab36g12.y1 Gm-c1026 Glycine max cDNA clone GEN
831	G1478	C95300	2.20E-10	[Citrus unshiu]	C95300 Citrus unshiu Miyagawa-wase maturation
831	G1478	AW034552	2.70E-10	[Lycopersicon esculentum]	EST278168 tomato callus, TAMU Lycop
831	G1478	BI070429	3.40E-10	[Populus tremula x Populus tremuloides]	C037P68U Populus tra
831	G1478	AF016011	5.10E-09	[Brassica napus]	CONSTANS homolog (Bn9CON10) gene, complete c
831	G1478	BE598912	6.20E-09	[Sorghum bicolor]	PI1_84_H11.b1_A002 Pathogen induced 1 (PI1)
831	G1478	BG605313	6.80E-09	[Triticum aestivum]	WHE2331_C04_F07ZS Wheat pre-anthesis spk
831	G1478	BE558327	8.90E-09	[Hordeum vulgare]	HV_CEB0017D19f Hordeum vulgare seedling gre
831	G1478	BG647091	1.20E-08	[Medicago truncatula]	EST508710 HOGA Medicago truncatula cDNA
831	G1478	gi2895188	4.70E-11	[Brassica napus]	CONSTANS homolog.
831	G1478	gi3618308	1.50E-09	[Oryza sativa]	zinc finger protein.
831	G1478	gi11037308	4.70E-09	[Brassica nigra]	constans-like protein.
831	G1478	gi3341723	1.30E-08	[Raphanus sativus]	CONSTANS-like 1 protein.
831	G1478	gi4091806	1.50E-07	[Malus x domestica]	CONSTANS-like protein 2.
831	G1478	gi10946337	3.10E-07	[Ipomoea nil]	CONSTANS-like protein.
831	G1478	gi4557093	1.40E-05	[Pinus radiata]	zinc finger protein.
831	G1478	gi619312	0.9	[Capparis masakaj]	mabinlin III B-chain=sweet protein mabi
831	G1478	gi4732091	1	[Zea mays]	bundle sheath defective protein 2.
831	G1478	gi4699629	1	[Nicotiana glauca]	Chain A, Putative Ancestral Protein Encod
579	G681	BG128147	6.80E-41	[Lycopersicon esculentum]	EST473793 tomato shoot/meristem Lyc
579	G681	BF054497	1.50E-39	[Solanum tuberosum]	EST439727 potato leaves and petioles Sola
579	G681	BE054276	8.40E-39	[Gossypium arboreum]	GA_Ea0002O18f Gossypium arboreum 7-10 d
579	G681	BG269414	4.00E-38	[Mesembryanthemum crystallinum]	L0-3478T3 Ice plant Lambda Un
579	G681	BF620286	7.40E-38	[Hordeum vulgare]	HVSMIEc0019F08f Hordeum vulgare seedling sho
579	G681	BE490032	1.00E-37	[Triticum aestivum]	WHE0364_C04_E08ZS Wheat cold-stressed see
579	G681	BI542536	1.40E-36	[Zea mays]	949021A03.y1 949 - Juvenile leaf and shoot cDNA fr
579	G681	BF425254	7.20E-36	[Glycine max]	su42c10.y1 Gm-c1068 Glycine max cDNA clone GENO
579	G681	AW672062	3.20E-34	[Sorghum bicolor]	LG1_354_G05.b1_A002 Light Grown 1 (LG1) Sor

Table 5

579	G681	BG448527	1.00E-33	[Medicago truncatula]	NF036F04RT1F1032 Developing root Medica
579	G681	gi13346188	9.10E-37	[Gossypium hirsutum]	GhMYB25.
579	G681	gi20563	6.30E-36	[Petunia x hybrida]	protein 1.
579	G681	gi485867	1.20E-34	[Antirrhinum majus]	mixta.
579	G681	gi2605617	1.70E-32	[Oryza sativa]	OSMYB1.
579	G681	gi1430846	2.00E-31	[Lycopersicon esculentum]	myb-related transcription factor.
579	G681	gi6651292	2.20E-30	[Pimpinella brachycarpa]	myb-related transcription factor.
579	G681	gi15042116	4.90E-30	[Zea mays subsp. parviglumis]	Cl protein.
579	G681	gi82730	6.10E-30	[Zea mays]	transforming protein (myb) homolog (clone Zm38)
579	G681	gi5139806	8.30E-30	[Glycine max]	GmMYB29A2.
579	G681	gi19055	1.10E-29	[Hordeum vulgare]	MybHv5.
611	G878	AF096299	6.20E-90	[Nicotiana tabacum]	DNA-binding protein 2 (WRKY2) mRNA, compl
611	G878	CUSSLD8	1.80E-83	[Cucumis sativus]	SPF1-like DNA-binding protein mRNA, complet
611	G878	AF193802	3.50E-63	[Oryza sativa]	zinc finger transcription factor WRKY1 mRNA, c
611	G878	AX192162	2.20E-62	[Glycine max]	Sequence 9 from Patent WO0149840.
611	G878	IPBSPF1P	3.80E-58	[Ipomoea batatas]	Sweet potato mRNA for SPF1 protein, complet
611	G878	AFABF1	2.00E-56	[Avena fatua]	A. fatua mRNA for DNA-binding protein (clone ABF
611	G878	LES303343	7.20E-55	[Lycopersicon esculentum]	mRNA for hypothetical protein (ORF
611	G878	AX192164	4.00E-54	[Triticum aestivum]	Sequence 11 from Patent WO0149840.
611	G878	AF080595	2.10E-53	[Pimpinella brachycarpa]	zinc finger protein (ZFP1) mRNA, com
611	G878	PCU48831	2.30E-53	[Petroselinum crispum]	DNA-binding protein WRKY1 mRNA, comple
611	G878	gi4322940	3.30E-128	[Nicotiana tabacum]	DNA-binding protein 2.
611	G878	gi927025	1.10E-109	[Cucumis sativus]	SPF1-like DNA-binding protein.
611	G878	gi6689916	1.50E-74	[Oryza sativa]	zinc finger transcription factor WRKY1.
611	G878	gi484261	1.10E-66	[Ipomoea batatas]	SPF1 protein.
611	G878	gi1159877	2.30E-63	[Avena fatua]	DNA-binding protein.
611	G878	gi13620227	4.60E-63	[Lycopersicon esculentum]	hypothetical protein.
611	G878	gi5917653	1.70E-56	[Petroselinum crispum]	zinc-finger type transcription facto
611	G878	gi4894965	5.00E-56	[Avena sativa]	DNA-binding protein WRKY1.
611	G878	gi3420906	8.70E-56	[Pimpinella brachycarpa]	zinc finger protein; WRKY1.
611	G878	gi13620168	4.20E-22	[Capsella rubella]	hypothetical protein.
47	G374	AP004457	1.20E-73	[Oryza sativa (japonica cultivar-group)]	() chromosome 8 clo
47	G374	AP004693	1.90E-73	[Oryza sativa]	chromosome 8 clone P0461F06, *** SEQUENCING IN
47	G374	BH552835	1.30E-62	[Brassica oleracea]	BOHGT56TR BOHG Brassica oleracea genomic
47	G374	BG128229	6.50E-55	[Lycopersicon esculentum]	EST473875 tomato shoot/meristem Lyc
47	G374	BG646959	3.20E-46	[Medicago truncatula]	EST508578 HOGA Medicago truncatula cDNA

Table 5

47	G374	BG890162	8.70E-41	[Solanum tuberosum]	EST516013 cSTD Solanum tuberosum cDNA clo
47	G374	AW179366	6.00E-38	[Zea mays]	618046G06.y1 618 - Inbred Tassel cDNA Library Zea
47	G374	BF473206	1.50E-32	[Triticum aestivum]	WHE0922 G12 M24ZS Wheat 5-15 DAP spike cD
47	G374	AW761011	2.90E-29	[Glycine max]	sl61g11.y1 Gm-c1027 Glycine max cDNA clone GENO
47	G374	AJ436050	1.50E-27	[Hordeum vulgare]	AJ436050 S00007 Hordeum vulgare cDNA clone
47	G374	gi422012	0.8	[Sorghum bicolor]	lipid transfer protein - sorghum (fragmen
47	G374	gi1827893	1	[Zea mays]	Maize Nonspecific Lipid Transfer Protein Complex

Traits of interest

Examples of some of the traits that may be desirable in plants, and that may be provided by transforming the plants with the presently disclosed sequences, are listed in Table 6.

Table 6. Genes, traits and utilities that affect plant characteristics

<u>Trait Category</u>	<u>Traits</u>	Transcription factor genes that <u>impact traits</u>	<u>Utility</u> <u>Gene effect on:</u>
Resistance and tolerance	Salt stress resistance	G22; G196; G226; G303; G312; G325; G353; G482; G545; G801; G867; G884; G922; G926; G1452; G1794; G1820; G1836; G1843; G1863; G2053; G2110; G2140; G2153; G2379; G2701; G2713; G2719; G2789	Germination rate, survivability, yield; extended growth range
	Osmotic stress resistance	G47; G175; G188; G303; G325; G353; G489; G502; G526; G921; G922; G926; G1069; G1089; G1452; G1794; G1930; G2140; G2153; G2379; G2701; G2719; G2789;	Germination rate, survivability, yield
	Cold stress resistance; cold germination	G256; G394; G664; G864; G1322; G2130	Germination, growth, earlier planting
	Tolerance to freezing	G303; G325; G353; G720; G912; G913; G1794; G2053; G2140; G2153; G2379; G2701; G2719; G2789	Survivability, yield, appearance, extended range
	Heat stress resistance	G3; G464; G682; G864; G964;	Germination,

		G1305; G1645; G2130 G2430	growth, later planting
	Drought, low humidity resistance	G303; G325; G353; G720; G912; G926; G1452; G1794; G1820; G1843; G2053; G2140; G2153; G2379; G2583; G2701; G2719; G2789	Survivability, yield, extended range
	Radiation resistance	G1052	Survivability, vigor, appearance
	Decreased herbicide sensitivity	G343; G2133; G2517	Resistant to increased herbicide use
	Increased herbicide sensitivity	G374; G877; G1519	Use as a herbicide target
	Oxidative stress	G477; G789; G1807; G2133; G2517	Improved yield, appearance, reduced senescence
	Light response	G183; G354; G375; G1062; G1322; G1331; G1488; G1494; G1521; G1786; G1794; G2144; G2555;	Germination, growth, development, flowering time
Development, morphology	Overall plant architecture	G24; G27; G31; G33; G47; G147; G156; G160; G182; G187; G195; G196; G211; G221; G237; G280; G342; G352; G357; G358; G360; G362; G364; G365; G367; G373; G377; G396; G431; G447; G479; G546; G546; G551; G578; G580; G596; G615; G617; G620; G625;	Vascular tissues, lignin content; cell wall content; appearance

		G638; G658; G716; G725; G727; G730; G740; G770; G858; G865; G869; G872; G904; G910; G912; G920; G939; G963; G977; G979; G987; G988; G993; G1007; G1010; G1014; G1035; G1046; G1049; G1062; G1069; G1070; G1076; G1089; G1093; G1127; G1131; G1145; G1229; G1246; G1304; G1318; G1320; G1330; G1331; G1352; G1354; G1360; G1364; G1379; G1384; G1399; G1415; G1417; G1442; G1453; G1454; G1459; G1460; G1471; G1475; G1477; G1487; G1487; G1492; G1499; G1499; G1531; G1540; G1543; G1543; G1544; G1548; G1584; G1587; G1588; G1589; G1636; G1642; G1747; G1749; G1749; G1751; G1752; G1763; G1766; G1767; G1778; G1789; G1790; G1791; G1793; G1794; G1795; G1800; G1806; G1811; G1835; G1836; G1838; G1839; G1843; G1853; G1855; G1865; G1881; G1882; G1883; G1884; G1891; G1896; G1898; G1902; G1904; G1906; G1913; G1914; G1925; G1929; G1930; G1954; G1958; G1965; G1976; G2057; G2107; G2133; G2134; G2151; G2154; G2157; G2181;	
--	--	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

		G2290; G2299; G2340; G2340; G2346; G2373; G2376; G2424; G2465; G2505; G2509; G2512; G2513; G2519; G2520; G2533; G2534; G2573; G2589; G2687; G2720; G2787; G2789; G2893	
	Size: increased stature	G189; G1073; G1435; G2430	
	Size: reduced stature or dwarfism	G3; G5; G21; G23; G39; G165; G184; G194; G258; G280; G340; G343; G353; G354; G362; G363; G370; G385; G396; G439; G440; G447; G450; G550; G557; G599; G636; G652; G670; G671; G674; G729; G760; G804; G831; G864; G884; G898; G900; G912; G913; G922; G932; G937; G939; G960; G962; G977; G991; G1000; G1008; G1020; G1023; G1053; G1067; G1075; G1137; G1181; G1198; G1228; G1266; G1267; G1275; G1277; G1309; G1311; G1314; G1317; G1322; G1323; G1326; G1332; G1334; G1367; G1381; G1382; G1386; G1421; G1488; G1494; G1537; G1545; G1560; G1586; G1641; G1652; G1655; G1671; G1750; G1756; G1757; G1782; G1786; G1794; G1839; G1845; G1879; G1886; G1888; G1933; G1939; G1943; G1944; G2011; G2094; G2115;	Ornamental; small stature provides wind resistance; creation of dwarf varieties

		G2130; G2132; G2144; G2145; G2147; G2156; G2294; G2313; G2344; G2431; G2510; G2517; G2521; G2893; G2893	
	Fruit size and number	G362	Biomass, yield, cotton boll fiber density
	Flower structure, inflorescence	G47; G259; G353; G354; G671; G732; G988; G1000; G1063; G1140; G1326; G1449; G1543; G1560; G1587; G1645; G1947; G2108; G2143; G2893	Ornamental horticulture; production of saffron or other edible flowers
	Number and development of trichomes	G225; G226; G247; G362; G585; G634; G676; G682; G1014; G1332; G1452; G1795; G2105	Resistance to pests and desiccation; essential oil production
	Seed size, color, and number	G156; G450; G584; G652; G668; G858; G979; G1040; G1062; G1145; G1255; G1494; G1531; G1534; G1594; G2105; G2114;	Yield
	Root development, modifications	G9; G1482; G1534; G1794; G1852; G2053; G2136; G2140	
	Modifications to root hairs	G225; G226	Nutrient, water uptake, pathogen resistance
	Apical dominance	G559; G732; G1255; G1275; G1411; G1488; G1635; G2452; G2509	Ornamental horticulture
	Branching patterns	G568; G988; G1548	Ornamental horticulture, knot reduction, improved

			windscreen
	Leaf shape, color, modifications	G375; G377; G428; G438; G447; G464; G557; G577; G599; G635; G671; G674; G736; G804; G903; G977; G921; G922; G1038; G1063; G1067; G1073; G1075; G1146; G1152; G1198; G1267; G1269; G1452; G1484; G1586; G1594; G1767; G1786; G1792; G1886; G2059; G2094; G2105; G2113; G2117; G2143; G2144; G2431; G2452; G2465; G2587; G2583; G2724;	Appealing shape or shiny leaves for ornamental agriculture, increased biomass or photosynthesis
	Silique	G1134	Ornamental
	Stem morphology	G47; G438; G671; G748; G988; G1000	Ornamental; digestibility
	Shoot modifications	G390; G391	Ornamental stem bifurcations
Disease, Pathogen Resistance	Bacterial	G211; G347; G367; G418; G525; G545; G578; G1049	Yield, appearance, survivability, extended range
	Fungal	G19; G28; G28; G28; G147; G188; G207; G211; G237; G248; G278; G347; G367; G371; G378; G409; G477; G545; G545; G558; G569; G578; G591; G594; G616; G789; G805; G812; G865; G869; G872; G881; G896; G940; G1047; G1049; G1064; G1084; G1196; G1255; G1266;	Yield, appearance, survivability, extended range

		G1363; G1514; G1756; G1792; G1792; G1792; G1792; G1880; G1919; G1919; G1927; G1927; G1936; G1936; G1950; G2069; G2130; G2380; G2380; G2555	
Nutrients	Increased tolerance to nitrogen-limited soils	G225; G226; G1792	
	Increased tolerance to phosphate-limited soils	G419; G545; G561; G1946	
	Increased tolerance to potassium-limited soils	G561; G911	
Hormonal	Hormone sensitivity	G12; G546; G926; G760; G913; G926; G1062; G1069; G1095; G1134; G1330; G1452; G1666; G1820; G2140; G2789	Seed dormancy, drought tolerance; plant form, fruit ripening
Seed biochemistry	Production of seed prenyl lipids, including tocopherol	G214; G259; G490; G652; G748; G883; G1052; G1328; G1930; G2509; G2520	Antioxidant activity, vitamin E
	Production of seed sterols	G20	Precursors for human steroid hormones; cholesterol modulators
	Production of seed glucosinolates	G353; G484; G674; G1272; G1506; G1897; G1946; G2113; G2117; G2155; G2290; G2340	Defense against insects; putative anticancer activity; undesirable in

			animal feeds
	Modified seed oil content	G162; G162; G180; G192; G241; G265; G286; G291; G427; G509; G519; G561; G567; G590; G818; G849; G892; G961; G974; G1063; G1143; G1190; G1198; G1226; G1229; G1323; G1451; G1471; G1478; G1496; G1526; G1543; G1640; G1644; G1646; G1672; G1677; G1750; G1765; G1777; G1793; G1838; G1902; G1946; G1948; G2059; G2123; G2138; G2139; G2343; G2792; G2830	Vegetable oil production; increased caloric value for animal feeds; lutein content
	Modified seed oil composition	G217; G504; G622; G778; G791; G861; G869; G938; G965; G1417; G2192	Heat stability, digestibility of seed oils
	Modified seed protein content	G162; G226; G241; G371; G427; G509; G567; G597; G732; G849; G865; G892; G963; G988; G1323; G1323; G1419; G1478; G1488; G1634; G1637; G1641; G1644; G1652; G1677; G1777; G1777; G1818; G1820; G1903; G1909; G1946; G1946; G1958; G2059; G2117; G2417; G2509	Reduced caloric value for humans
Leaf biochemistry	Production of flavonoids	G1666*	Ornamental pigment production; pathogen resistance; health

			benefits
	Production of leaf glucosinolates	G264; G353; G484; G652; G674; G681; G1069; G1198; G1322; G1421; G1657; G1794; G1897; G1946; G2115; G2117; G2144; G2155; G2155; G2340; G2512; G2520; G2552	Defense against insects; putative anticancer activity; undesirable in animal feeds
	Production of diterpenes	G229	Induction of enzymes involved in alkaloid biosynthesis
	Production of anthocyanin	G546	Ornamental pigment
	Production of leaf phytosterols, inc. stigmastanol, campesterol	G561; G2131; G2424	Precursors for human steroid hormones; cholesterol modulators
	Leaf fatty acid composition	G214; G377; G861; G962; G975; G987; G1266; G1337; G1399; G1465; G1512; G2136; G2147; G2192	Nutritional value; increase in waxes for disease resistance
	Production of leaf prenyl lipids, including tocopherol	G214; G259; G280; G652; G987; G1543; G2509; G2520	Antioxidant activity, vitamin E
Biochemistry, general	Production of miscellaneous secondary metabolites	G229; G663	
	Sugar, starch, hemicellulose composition,	G158; G211; G211; G237; G242; G274; G598; G1012; G1266; G1309; G1309; G1641; G1765; G1865; G2094; G2094;	Food digestibility, hemicellulose & pectin content; fiber content; plant

		G2589; G2589	tensile strength, wood quality, pathogen resistance, pulp production; tuber starch content
Sugar sensing	Plant response to sugars	G26; G38; G43; G207; G218; G241; G254; G263; G308; G536; G567; G567; G680; G867; G912; G956; G996; G1068; G1225; G1314; G1314; G1337; G1759; G1804; G2153; G2379	Photosynthetic rate, carbohydrate accumulation, biomass production, source-sink relationships, senescence
Growth, Reproduction	Plant growth rate and development	G447; G617; G674; G730; G917; G937; G1035; G1046; G1131; G1425; G1452; G1459; G1492; G1589; G1652; G1879; G1943; G2430; G2431; G2465; G2521	Faster growth, increased biomass or yield, improved appearance; delay in bolting
	Embryo development	G167	
	Seed germination rate	G979; G1792; G2130	Yield
	Plant, seedling vigor	G561; G2346	Survivability, yield
	Senescence; cell death	G571; G636; G878; G1050; G1463; G1749; G1944; G2130; G2155; G2340; G2383	Yield, appearance; response to pathogens;
	Modified fertility	G39; G340; G439; G470; G559; G615; G652; G671; G779; G962; G977; G988; G1000; G1063; G1067; G1075;	Prevents or minimizes escape of the pollen of GMOs

		G1266; G1311; G1321; G1326; G1367; G1386; G1421; G1453; G1471; G1453; G1560; G1594; G1635; G1750; G1947; G2011; G2094; G2113; G2115; G2130; G2143; G2147; G2294; G2510; G2893	
	Early flowering	G147; G157; G180; G183; G183; G184; G185; G208; G227; G294; G390; G390; G390; G391; G391; G427; G427; G490; G565; G590; G592; G720; G789; G865; G898; G898; G989; G989; G1037; G1037; G1142; G1225; G1225; G1226; G1242; G1305; G1305; G1380; G1380; G1480; G1480; G1488; G1494; G1545; G1545; G1649; G1706; G1760; G1767; G1767; G1820; G1841; G1841; G1842; G1843; G1843; G1946; G1946; G2010; G2030; G2030; G2144; G2144; G2295; G2295; G2347; G2348; G2348; G2373; G2373; G2509; G2509; G2555; G2555	Faster generation time; synchrony of flowering; potential for introducing new traits to single variety
	Delayed flowering	G8; G47; G192; G214; G234; G361; G362; G562; G568; G571; G591; G680; G736; G748; G859; G878; G910; G912; G913; G971; G994; G1051; G1052; G1073; G1079; G1335; G1435; G1452; G1478;	Delayed time to pollen production of GMO plants; synchrony of flowering; increased yield

		G1789; G1804; G1865; G1865; G1895; G1900; G2007; G2133; G2155; G2291; G2465	
	Extended flowering phase	G1947	
	Flower and leaf development	G259; G353; G377; G580; G638 G652; G858; G869; G917; G922; G932; G1063; G1075; G1140; G1425; G1452; G1499; G1548; G1645; G1865; G1897; G1933; G2094; G2124; G2140; G2143; G2535; G2557	Ornamental applications; decreased fertility
	Flower abscission	G1897	Ornamental: longer retention of flowers

* When co-expressed with G669 and G663

Significance of modified plant traits

Currently, the existence of a series of maturity groups for different latitudes represents a major barrier to the introduction of new valuable traits. Any trait (e.g. disease resistance) has to be bred into each of the different maturity groups separately, a laborious and costly exercise. The availability of single strain, which could be grown at any latitude, would therefore greatly increase the potential for introducing new traits to crop species such as soybean and cotton.

For many of the traits, listed in Table 6 and below, that may be conferred to plants, a single transcription factor gene may be used to increase or decrease, advance or delay, or improve or prove deleterious to a given trait. For example, overexpression of a transcription factor gene that naturally occurs in a plant may cause early flowering relative to non-transformed or wild-type plants. By knocking out the gene, or suppressing the gene (with, for example, antisense suppression) the plant may experience delayed flowering. Similarly, overexpressing or suppressing one or more genes can impart significant differences in production of plant products,

such as different fatty acid ratios. Thus, suppressing a gene that causes a plant to be more sensitive to cold may improve a plant's tolerance of cold.

Salt stress resistance. Soil salinity is one of the more important variables that determines where a plant may thrive. Salinity is especially important for the successful cultivation of crop plants, particular in many parts of the world that have naturally high soil salt concentrations, or where the soil has been over-utilized. Thus, presently disclosed transcription factor genes that provide increased salt tolerance during germination, the seedling stage, and throughout a plant's life cycle would find particular value for imparting survivability and yield in areas where a particular crop would not normally prosper.

Osmotic stress resistance. Presently disclosed transcription factor genes that confer resistance to osmotic stress may increase germination rate under adverse conditions, which could impact survivability and yield of seeds and plants.

Cold stress resistance. The potential utility of presently disclosed transcription factor genes that increase tolerance to cold is to confer better germination and growth in cold conditions. The germination of many crops is very sensitive to cold temperatures. Genes that would allow germination and seedling vigor in the cold would have highly significant utility in allowing seeds to be planted earlier in the season with a high rate of survivability. Transcription factor genes that confer better survivability in cooler climates allow a grower to move up planting time in the spring and extend the growing season further into autumn for higher crop yields.

Tolerance to freezing. The presently disclosed transcription factor genes that impart tolerance to freezing conditions are useful for enhancing the survivability and appearance of plants conditions or conditions that would otherwise cause extensive cellular damage. Thus, germination of seeds and survival may take place at temperatures significantly below that of the mean temperature required for germination of seeds and survival of non-transformed plants. As with salt tolerance, this has the added benefit of increasing the potential range of a crop plant into regions in which it would otherwise succumb. Cold tolerant transformed plants may also be

planted earlier in the spring or later in autumn, with greater success than with non-transformed plants.

Heat stress tolerance. The germination of many crops is also sensitive to high temperatures. Presently disclosed transcription factor genes that provide increased heat tolerance are generally useful in producing plants that germinate and grow in hot conditions, may find particular use for crops that are planted late in the season, or extend the range of a plant by allowing growth in relatively hot climates.

Drought, low humidity tolerance. Strategies that allow plants to survive in low water conditions may include, for example, reduced surface area or surface oil or wax production. A number of presently disclosed transcription factor genes increase a plant's tolerance to low water conditions and provide the benefits of improved survivability, increased yield and an extended geographic and temporal planting range.

Radiation resistance. Presently disclosed transcription factor genes have been shown to increase lutein production. Lutein, like other xanthophylls such as zeaxanthin and violaxanthin, are important in the protection of plants against the damaging effects of excessive light. Lutein contributes, directly or indirectly, to the rapid rise of non-photochemical quenching in plants exposed to high light. Increased tolerance of field plants to visible and ultraviolet light impacts survivability and vigor, particularly for recent transplants. Also affected are the yield and appearance of harvested plants or plant parts. Crop plants engineered with presently disclosed transcription factor genes that cause the plant to produce higher levels of lutein therefore would have improved photoprotection, leading to less oxidative damage and increase vigor, survivability and higher yields under high light and ultraviolet light conditions.

Decreased herbicide sensitivity. Presently disclosed transcription factor genes that confer resistance or tolerance to herbicides (e.g., glyphosate) may find use in providing means to increase herbicide applications without detriment to desirable plants. This would allow for the increased use of a particular herbicide in a local

environment, with the effect of increased detriment to undesirable species and less harm to transgenic, desirable cultivars.

Increased herbicide sensitivity. Knockouts of a number of the presently disclosed transcription factor genes have been shown to be lethal to developing embryos. Thus, these genes are potentially useful as herbicide targets.

Oxidative stress. In plants, as in all living things, abiotic and biotic stresses induce the formation of oxygen radicals, including superoxide and peroxide radicals. This has the effect of accelerating senescence, particularly in leaves, with the resulting loss of yield and adverse effect on appearance. Generally, plants that have the highest level of defense mechanisms, such as, for example, polyunsaturated moieties of membrane lipids, are most likely to thrive under conditions that introduce oxidative stress (e.g., high light, ozone, water deficit, particularly in combination). Introduction of the presently disclosed transcription factor genes that increase the level of oxidative stress defense mechanisms would provide beneficial effects on the yield and appearance of plants. One specific oxidizing agent, ozone, has been shown to cause significant foliar injury, which impacts yield and appearance of crop and ornamental plants. In addition to reduced foliar injury that would be found in ozone resistant plant created by transforming plants with some of the presently disclosed transcription factor genes, the latter have also been shown to have increased chlorophyll fluorescence (Yu-Sen Chang et al. Bot. Bull. Acad. Sin. (2001) 42: 265-272).

Heavy metal tolerance. Heavy metals such as lead, mercury, arsenic, chromium and others may have a significant adverse impact on plant respiration. Plants that have been transformed with presently disclosed transcription factor genes that confer improved resistance to heavy metals, through, for example, sequestering or reduced uptake of the metals will show improved vigor and yield in soils with relatively high concentrations of these elements. Conversely, transgenic transcription factors may also be introduced into plants to confer an increase in heavy metal uptake, which may benefit efforts to clean up contaminated soils.

Light response. Presently disclosed transcription factor genes that modify a plant's response to light may be useful for modifying a plant's growth or

development, for example, photomorphogenesis in poor light, or accelerating flowering time in response to various light intensities, quality or duration to which a non-transformed plant would not similarly respond. Examples of such responses that have been demonstrated include leaf number and arrangement, and early flower bud appearances.

Overall plant architecture. Several presently disclosed transcription factor genes have been introduced into plants to alter numerous aspects of the plant's morphology. For example, it has been demonstrated that a number of transcription factors may be used to manipulate branching, such as the means to modify lateral branching, a possible application in the forestry industry. Transgenic plants have also been produced that have altered cell wall content, lignin production, flower organ number, or overall shape of the plants. Presently disclosed transcription factor genes transformed into plants may be used to affect plant morphology by increasing or decreasing internode distance, both of which may be advantageous under different circumstances. For example, for fast growth of woody plants to provide more biomass, or fewer knots, increased internode distances are generally desirable. For improved wind screening of shrubs or trees, or harvesting characteristics of, for example, members of the Gramineae family, decreased internode distance may be advantageous. These modifications would also prove useful in the ornamental horticulture industry for the creation of unique phenotypic characteristics of ornamental plants.

Increased stature. For some ornamental plants, the ability to provide larger varieties may be highly desirable. For many plants, including fruit-bearing trees or trees and shrubs that serve as view or wind screens, increased stature provides obvious benefits. Crop species may also produce higher yields on larger cultivars.

Reduced stature or dwarfism. Presently disclosed transcription factor genes that decrease plant stature can be used to produce plants that are more resistant to damage by wind and rain, or more resistant to heat or low humidity or water deficit. Dwarf plants are also of significant interest to the ornamental horticulture industry, and particularly for home garden applications for which space availability may be limited.

Fruit size and number. Introduction of presently disclosed transcription factor genes that affect fruit size will have desirable impacts on fruit size and number, which may comprise increases in yield for fruit crops, or reduced fruit yield, such as when vegetative growth is preferred (e.g., with bushy ornamentals, or where fruit is undesirable, as with ornamental olive trees).

Flower structure, inflorescence, and development. Presently disclosed transgenic transcription factors have been used to create plants with larger flowers or arrangements of flowers that are distinct from wild-type or non-transformed cultivars. This would likely have the most value for the ornamental horticulture industry, where larger flowers or interesting presentations generally are preferred and command the highest prices. Flower structure may have advantageous effects on fertility, and could be used, for example, to decrease fertility by the absence, reduction or screening of reproductive components. One interesting application for manipulation of flower structure, for example, by introduced transcription factors could be in the increased production of edible flowers or flower parts, including saffron, which is derived from the stigmas of *Crocus sativus*.

Number and development of trichomes. Several presently disclosed transcription factor genes have been used to modify trichome number and amount of trichome products in plants. Trichome glands on the surface of many higher plants produce and secrete exudates that give protection from the elements and pests such as insects, microbes and herbivores. These exudates may physically immobilize insects and spores, may be insecticidal or ant-microbial or they may act as allergens or irritants to protect against herbivores. Trichomes have also been suggested to decrease transpiration by decreasing leaf surface air flow, and by exuding chemicals that protect the leaf from the sun.

Seed size, color and number. The introduction of presently disclosed transcription factor genes into plants that alter the size or number of seeds may have a significant impact on yield, both when the product is the seed itself, or when biomass of the vegetative portion of the plant is increased by reducing seed production. In the case of fruit products, it is often advantageous to modify a plant to have reduced size

or number of seeds relative to non-transformed plants to provide seedless or varieties with reduced numbers or smaller seeds. Presently disclosed transcription factor genes have also been shown to affect seed size, including the development of larger seeds. Seed size, in addition to seed coat integrity, thickness and permeability, seed water content and by a number of other components including antioxidants and oligosaccharides, may affect seed longevity in storage. This would be an important utility when the seed of a plant is the harvested crops, as with, for example, peas, beans, nuts, etc. Presently disclosed transcription factor genes have also been used to modify seed color, which could provide added appeal to a seed product.

Root development, modifications. By modifying the structure or development of roots by transforming into a plant one or more of the presently disclosed transcription factor genes, plants may be produced that have the capacity to thrive in otherwise unproductive soils. For example, grape roots that extend further into rocky soils, or that remain viable in waterlogged soils, would increase the effective planting range of the crop. It may be advantageous to manipulate a plant to produce short roots, as when a soil in which the plant will be growing is occasionally flooded, or when pathogenic fungi or disease-causing nematodes are prevalent.

Modifications to root hairs. Presently disclosed transcription factor genes that increase root hair length or number potentially could be used to increase root growth or vigor, which might in turn allow better plant growth under adverse conditions such as limited nutrient or water availability.

Apical dominance. The modified expression of presently disclosed transcription factors that control apical dominance could be used in ornamental horticulture, for example, to modify plant architecture.

Branching patterns. Several presently disclosed transcription factor genes have been used to manipulate branching, which could provide benefits in the forestry industry. For example, reduction in the formation of lateral branches could reduce knot formation. Conversely, increasing the number of lateral branches could provide utility when a plant is used as a windscreen, or may also provide ornamental advantages.

Leaf shape, color and modifications. It has been demonstrated in laboratory experiments that overexpression of some of the presently disclosed transcription factors produced marked effects on leaf development. At early stages of growth, these transgenic seedlings developed narrow, upward pointing leaves with long petioles, possibly indicating a disruption in circadian-clock controlled processes or nyctinastic movements. Other transcription factor genes can be used to increase plant biomass; large size would be useful in crops where the vegetative portion of the plant is the marketable portion.

Siliques. Genes that later silique conformation in brassicates may be used to modify fruit ripening processes in brassicates and other plants, which may positively affect seed or fruit quality.

Stem morphology and shoot modifications. Laboratory studies have demonstrated that introducing several of the presently disclosed transcription factor genes into plants can cause stem bifurcations in shoots, in which the shoot meristems split to form two or three separate shoots. This unique appearance would be desirable in ornamental applications.

Diseases, pathogens and pests. A number of the presently disclosed transcription factor genes have been shown to or are likely to confer resistance to various plant diseases, pathogens and pests. The offending organisms include fungal pathogens *Fusarium oxysporum*, *Botrytis cinerea*, *Sclerotinia sclerotiorum*, and *Erysiphe orontii*. Bacterial pathogens to which resistance may be conferred include *Pseudomonas syringae*. Other problem organisms may potentially include nematodes, mollicutes, parasites, or herbivorous arthropods. In each case, one or more transformed transcription factor genes may provide some benefit to the plant to help prevent or overcome infestation. The mechanisms by which the transcription factors work could include increasing surface waxes or oils, surface thickness, local senescence, or the activation of signal transduction pathways that regulate plant defense in response to attacks by herbivorous pests (including, for example, protease inhibitors).

Increased tolerance of plants to nutrient-limited soils. Presently disclosed transcription factor genes introduced into plants may provide the means to improve uptake of essential nutrients, including nitrogenous compounds, phosphates, potassium, and trace minerals. The effect of these modifications is to increase the seedling germination and range of ornamental and crop plants. The utilities of presently disclosed transcription factor genes conferring tolerance to conditions of low nutrients also include cost savings to the grower by reducing the amounts of fertilizer needed, environmental benefits of reduced fertilizer runoff; and improved yield and stress tolerance. In addition, this gene could be used to alter seed protein amounts and/or composition that could impact yield as well as the nutritional value and production of various food products.

Hormone sensitivity. One or more of the presently disclosed transcription factor genes have been shown to affect plant abscisic acid (ABA) sensitivity. This plant hormone is likely the most important hormone in mediating the adaptation of a plant to stress. For example, ABA mediates conversion of apical meristems into dormant buds. In response to increasingly cold conditions, the newly developing leaves growing above the meristem become converted into stiff bud scales that closely wrap the meristem and protect it from mechanical damage during winter. ABA in the bud also enforces dormancy; during premature warm spells, the buds are inhibited from sprouting. Bud dormancy is eliminated after either a prolonged cold period of cold or a significant number of lengthening days. Thus, by affecting ABA sensitivity, introduced transcription factor genes may affect cold sensitivity and survivability. ABA is also important in protecting plants from drought tolerance.

Several other of the present transcription factor genes have been used to manipulate ethylene signal transduction and response pathways. These genes can thus be used to manipulate the processes influenced by ethylene, such as seed germination or fruit ripening, and to improve seed or fruit quality.

Production of seed and leaf prenyl lipids, including tocopherol. Prenyl lipids play a role in anchoring proteins in membranes or membranous organelles. Thus modifying the prenyl lipid content of seeds and leaves could affect membrane integrity and function. A number of presently disclosed transcription factor genes

have been shown to modify the tocopherol composition of plants. Tocopherols have both anti-oxidant and vitamin E activity.

Production of seed and leaf phytosterols: Presently disclosed transcription factor genes that modify levels of phytosterols in plants may have at least two utilities. First, phytosterols are an important source of precursors for the manufacture of human steroid hormones. Thus, regulation of transcription factor expression or activity could lead to elevated levels of important human steroid precursors for steroid semi-synthesis. For example, transcription factors that cause elevated levels of campesterol in leaves, or sitosterols and stigmasterols in seed crops, would be useful for this purpose. Phytosterols and their hydrogenated derivatives phytostanols also have proven cholesterol-lowering properties, and transcription factor genes that modify the expression of these compounds in plants would thus provide health benefits.

Production of seed and leaf glucosinolates: Some glucosinolates have anti-cancer activity; thus, increasing the levels or composition of these compounds by introducing several of the presently disclosed transcription factors might be of interest from a nutraceutical standpoint. (3) Glucosinolates form part of a plants natural defense against insects. Modification of glucosinolate composition or quantity could therefore afford increased protection from predators. Furthermore, in edible crops, tissue specific promoters might be used to ensure that these compounds accumulate specifically in tissues, such as the epidermis, which are not taken for consumption.

Modified seed oil content: The composition of seeds, particularly with respect to seed oil amounts and/or composition, is very important for the nutritional value and production of various food and feed products. Several of the presently disclosed transcription factor genes in seed lipid saturation that alter seed oil content could be used to improve the heat stability of oils or to improve the nutritional quality of seed oil, by, for example, reducing the number of calories in seed, increasing the number of calories in animal feeds, or altering the ratio of saturated to unsaturated lipids comprising the oils.

Seed and leaf fatty acid composition. A number of the presently disclosed transcription factor genes have been shown to alter the fatty acid composition in plants, and seeds in particular. This modification may find particular value for improving the nutritional value of, for example, seeds or whole plants. Dietary fatty acids ratios have been shown to have an effect on, for example, bone integrity and remodeling (see, for example, Weiler, H.A., *Pediatr Res* (2000) 47:5 692-697). The ratio of dietary fatty acids may alter the precursor pools of long-chain polyunsaturated fatty acids that serve as precursors for prostaglandin synthesis. In mammalian connective tissue, prostaglandins serve as important signals regulating the balance between resorption and formation in bone and cartilage. Thus dietary fatty acid ratios altered in seeds may affect the etiology and outcome of bone loss.

Modified seed protein content. As with seed oils, the composition of seeds, particularly with respect to protein amounts and/or composition, is very important for the nutritional value and production of various food and feed products. A number of the presently disclosed transcription factor genes modify the protein concentrations in seeds would provide nutritional benefits, and may be used to prolong storage, increase seed pest or disease resistance, or modify germination rates.

Production of flavonoids in leaves and other plant parts. Expression of presently disclosed transcription factor genes that increase flavonoid production in plants, including anthocyanins and condensed tannins, may be used to alter in pigment production for horticultural purposes, and possibly increasing stress resistance. Flavonoids have antimicrobial activity and could be used to engineer pathogen resistance. Several flavonoid compounds have health promoting effects such as the inhibition of tumor growth and cancer, prevention of bone loss and the prevention of the oxidation of lipids. Increasing levels of condensed tannins, whose biosynthetic pathway is shared with anthocyanin biosynthesis, in forage legumes is an important agronomic trait because they prevent pasture bloat by collapsing protein foams within the rumen. For a review on the utilities of flavonoids and their derivatives, refer to Dixon et al. (1999) *Trends Plant Sci.* 4:394-400.

Production of diterpenes in leaves and other plant parts. Depending on the plant species, varying amounts of diverse secondary biochemicals (often lipophilic

terpenes) are produced and exuded or volatilized by trichomes. These exotic secondary biochemicals, which are relatively easy to extract because they are on the surface of the leaf, have been widely used in such products as flavors and aromas, drugs, pesticides and cosmetics. Thus, the overexpression of genes that are used to produce diterpenes in plants may be accomplished by introducing transcription factor genes that induce said overexpression. One class of secondary metabolites, the diterpenes, can effect several biological systems such as tumor progression, prostaglandin synthesis and tissue inflammation. In addition, diterpenes can act as insect pheromones, termite allomones, and can exhibit neurotoxic, cytotoxic and antimitotic activities. As a result of this functional diversity, diterpenes have been the target of research several pharmaceutical ventures. In most cases where the metabolic pathways are impossible to engineer, increasing trichome density or size on leaves may be the only way to increase plant productivity.

Production of anthocyanin in leaves and other plant parts. Several presently disclosed transcription factor genes can be used to alter anthocyanin production in numerous plant species. The potential utilities of these genes include alterations in pigment production for horticultural purposes, and possibly increasing stress resistance in combination with another transcription factor.

Production of miscellaneous secondary metabolites. Microarray data suggests that flux through the aromatic amino acid biosynthetic pathways and primary and secondary metabolite biosynthetic pathways are up-regulated. Presently disclosed transcription factors have been shown to be involved in regulating alkaloid biosynthesis, in part by up-regulating the enzymes indole-3-glycerol phosphatase and strictosidine synthase. Phenylalanine ammonia lyase, chalcone synthase and trans-cinnamate mono-oxygenase are also induced, and are involved in phenylpropanoid biosynthesis.

Sugar, starch, hemicellulose composition. Overexpression of the presently disclosed transcription factors that affect sugar content resulted in plants with altered leaf insoluble sugar content. Transcription factors that alter plant cell wall composition have several potential applications including altering food digestibility, plant tensile strength, wood quality, pathogen resistance and in pulp production. The

potential utilities of a gene involved in glucose-specific sugar sensing are to alter energy balance, photosynthetic rate, carbohydrate accumulation, biomass production, source-sink relationships, and senescence.

Hemicellulose is not desirable in paper pulps because of its lack of strength compared with cellulose. Thus modulating the amounts of cellulose vs. hemicellulose in the plant cell wall is desirable for the paper/lumber industry. Increasing the insoluble carbohydrate content in various fruits, vegetables, and other edible consumer products will result in enhanced fiber content. Increased fiber content would not only provide health benefits in food products, but might also increase digestibility of forage crops. In addition, the hemicellulose and pectin content of fruits and berries affects the quality of jam and catsup made from them. Changes in hemicellulose and pectin content could result in a superior consumer product.

Plant response to sugars and sugar composition. In addition to their important role as an energy source and structural component of the plant cell, sugars are central regulatory molecules that control several aspects of plant physiology, metabolism and development. It is thought that this control is achieved by regulating gene expression and, in higher plants, sugars have been shown to repress or activate plant genes involved in many essential processes such as photosynthesis, glyoxylate metabolism, respiration, starch and sucrose synthesis and degradation, pathogen response, wounding response, cell cycle regulation, pigmentation, flowering and senescence. The mechanisms by which sugars control gene expression are not understood.

Because sugars are important signaling molecules, the ability to control either the concentration of a signaling sugar or how the plant perceives or responds to a signaling sugar could be used to control plant development, physiology or metabolism. For example, the flux of sucrose (a disaccharide sugar used for systemically transporting carbon and energy in most plants) has been shown to affect gene expression and alter storage compound accumulation in seeds. Manipulation of the sucrose signaling pathway in seeds may therefore cause seeds to have more protein, oil or carbohydrate, depending on the type of manipulation. Similarly, in tubers, sucrose is converted to starch which is used as an energy store. It is thought that sugar signaling pathways may partially determine the levels of starch synthesized

in the tubers. The manipulation of sugar signaling in tubers could lead to tubers with a higher starch content.

Thus, the presently disclosed transcription factor genes that manipulate the sugar signal transduction pathway may lead to altered gene expression to produce plants with desirable traits. In particular, manipulation of sugar signal transduction pathways could be used to alter source-sink relationships in seeds, tubers, roots and other storage organs leading to increase in yield.

Plant growth rate and development. A number of the presently disclosed transcription factor genes have been shown to have significant effects on plant growth rate and development. These observations have included, for example, more rapid or delayed growth and development of reproductive organs. This would provide utility for regions with short or long growing seasons, respectively. Accelerating plant growth would also improve early yield or increase biomass at an earlier stage, when such is desirable (for example, in producing forestry products).

Embryo development. Presently disclosed transcription factor genes that alter embryo development has been used to alter seed protein and oil amounts and/or composition which is very important for the nutritional value and production of various food products. Seed shape and seed coat may also be altered by these genes, which may provide for improved storage stability.

Seed germination rate. A number of the presently disclosed transcription factor genes have been shown to modify seed germination rate, including when the seeds are in conditions normally unfavorable for germination (e.g., cold, heat or salt stress, or in the presence of ABA), and may thus be used to modify and improve germination rates under adverse conditions.

Plant, seedling vigor. Seedlings transformed with presently disclosed transcription factors have been shown to possess larger cotyledons and appeared somewhat more advanced than control plants. This indicates that the seedlings developed more rapidly than the control plants. Rapid seedling development is likely to reduce loss due to diseases particularly prevalent at the seedling stage (e.g.,

damping off) and is thus important for survivability of plants germinating in the field or in controlled environments.

Senescence, cell death. Presently disclosed transcription factor genes may be used to alter senescence responses in plants. Although leaf senescence is thought to be an evolutionary adaptation to recycle nutrients, the ability to control senescence in an agricultural setting has significant value. For example, a delay in leaf senescence in some maize hybrids is associated with a significant increase in yields and a delay of a few days in the senescence of soybean plants can have a large impact on yield. Delayed flower senescence may also generate plants that retain their blossoms longer and this may be of potential interest to the ornamental horticulture industry.

Modified fertility. Plants that overexpress a number of the presently disclosed transcription factor genes have been shown to possess reduced fertility. This could be a desirable trait, as it could be exploited to prevent or minimize the escape of the pollen of genetically modified organisms (GMOs) into the environment.

Early and delayed flowering. Presently disclosed transcription factor genes that accelerate flowering could have valuable applications in such programs since they allow much faster generation times. In a number of species, for example, broccoli, cauliflower, where the reproductive parts of the plants constitute the crop and the vegetative tissues are discarded, it would be advantageous to accelerate time to flowering. Accelerating flowering could shorten crop and tree breeding programs. Additionally, in some instances, a faster generation time might allow additional harvests of a crop to be made within a given growing season. A number of *Arabidopsis* genes have already been shown to accelerate flowering when constitutively expressed. These include LEAFY, APETALA1 and CONSTANS (Mandel, M. et al., 1995, Nature 377, 522-524; Weigel, D. and Nilsson, O., 1995, Nature 377, 495-500; Simon et al., 1996, Nature 384, 59-62).

By regulating the expression of potential flowering using inducible promoters, flowering could be triggered by application of an inducer chemical. This would allow flowering to be synchronized across a crop and facilitate more efficient harvesting. Such inducible systems could also be used to tune the flowering of crop varieties to

different latitudes. At present, species such as soybean and cotton are available as a series of maturity groups that are suitable for different latitudes on the basis of their flowering time (which is governed by day-length). A system in which flowering could be chemically controlled would allow a single high-yielding northern maturity group to be grown at any latitude. In southern regions such plants could be grown for longer, thereby increasing yields, before flowering was induced. In more northern areas, the induction would be used to ensure that the crop flowers prior to the first winter frosts.

In a sizeable number of species, for example, root crops, where the vegetative parts of the plants constitute the crop and the reproductive tissues are discarded, it would be advantageous to delay or prevent flowering. Extending vegetative development with presently disclosed transcription factor genes could thus bring about large increases in yields.. Prevention of flowering might help maximize vegetative yields and prevent escape of genetically modified organism (GMO) pollen.

Extended flowering phase. Presently disclosed transcription factors that extend flowering time have utility in engineering plants with longer-lasting flowers for the horticulture industry, and for extending the time in which the plant is fertile.

Flower and leaf development. Presently disclosed transcription factor genes have been used to modify the development of flowers and leaves. This could be advantageous in the development of new ornamental cultivars that present unique configurations. In addition, some of these genes have been shown to reduce a plant's fertility, which is also useful for helping to prevent development of pollen of GMOs.

Flower abscission. Presently disclosed transcription factor genes introduced into plants have been used to retain flowers for longer periods. This would provide a significant benefit to the ornamental industry, for both cut flowers and woody plant varieties (of, for example, maize), as well as have the potential to lengthen the fertile period of a plant, which could positively impact yield and breeding programs.

A listing of specific effects and utilities that the presently disclosed transcription factor genes have on plants, as determined by direct observation and assay analysis, is provided in Table 4.

XVI. Antisense and Co-suppression

In addition to expression of the nucleic acids of the invention as gene replacement or plant phenotype modification nucleic acids, the nucleic acids are also useful for sense and anti-sense suppression of expression, e.g., to down-regulate expression of a nucleic acid of the invention, e.g., as a further mechanism for modulating plant phenotype. That is, the nucleic acids of the invention, or subsequences or anti-sense sequences thereof, can be used to block expression of naturally occurring homologous nucleic acids. A variety of sense and anti-sense technologies are known in the art, e.g., as set forth in Lichtenstein and Nellen (1997) Antisense Technology: A Practical Approach IRL Press at Oxford University Press, Oxford, U.K.. In general, sense or anti-sense sequences are introduced into a cell, where they are optionally amplified, e.g., by transcription. Such sequences include both simple oligonucleotide sequences and catalytic sequences such as ribozymes.

For example, a reduction or elimination of expression (i.e., a “knock-out”) of a transcription factor or transcription factor homologue polypeptide in a transgenic plant, e.g., to modify a plant trait, can be obtained by introducing an antisense construct corresponding to the polypeptide of interest as a cDNA. For antisense suppression, the transcription factor or homologue cDNA is arranged in reverse orientation (with respect to the coding sequence) relative to the promoter sequence in the expression vector. The introduced sequence need not be the full length cDNA or gene, and need not be identical to the cDNA or gene found in the plant type to be transformed. Typically, the antisense sequence need only be capable of hybridizing to the target gene or RNA of interest. Thus, where the introduced sequence is of shorter length, a higher degree of homology to the endogenous transcription factor sequence will be needed for effective antisense suppression. While antisense sequences of various lengths can be utilized, preferably, the introduced antisense sequence in the vector will be at least 30 nucleotides in length, and improved antisense suppression will typically be observed as the length of the antisense sequence increases. Preferably, the length of the antisense sequence in the vector will be greater than 100 nucleotides.

Transcription of an antisense construct as described results in the production of RNA molecules that are the reverse complement of mRNA molecules transcribed from the endogenous transcription factor gene in the plant cell.

Suppression of endogenous transcription factor gene expression can also be achieved using a ribozyme. Ribozymes are RNA molecules that possess highly specific endoribonuclease activity. The production and use of ribozymes are disclosed in U.S. Patent No. 4,987,071 and U.S. Patent No. 5,543,508. Synthetic ribozyme sequences including antisense RNAs can be used to confer RNA cleaving activity on the antisense RNA, such that endogenous mRNA molecules that hybridize to the antisense RNA are cleaved, which in turn leads to an enhanced antisense inhibition of endogenous gene expression.

Suppression of endogenous transcription factor gene expression can also be achieved using RNA interference, or RNAi. RNAi is a post-transcriptional, targeted gene-silencing technique that uses double-stranded RNA (dsRNA) to incite degradation of messenger RNA (mRNA) containing the same sequence as the dsRNA (Constans, (2002) *The Scientist* 16:36). Small interfering RNAs, or siRNAs are produced in at least two steps: an endogenous ribonuclease cleaves longer dsRNA into shorter, 21-23 nucleotide-long RNAs. The siRNA segments then mediate the degradation of the target mRNA (Zamore, (2001) *Nature Struct. Biol.*, 8:746-50). RNAi has been used for gene function determination in a manner similar to antisense oligonucleotides (Constans, (2002) *The Scientist* 16:36). Expression vectors that continually express siRNAs in transiently and stably transfected have been engineered to express small hairpin RNAs (shRNAs), which get processed in vivo into siRNA-like molecules capable of carrying out gene-specific silencing (Brummelkamp et al., (2002) *Science* 296:550-553, and Paddison, et al. (2002) *Genes & Dev.* 16:948-958). Post-transcriptional gene silencing by double-stranded RNA is discussed in further detail by Hammond et al. (2001) *Nature Rev Gen* 2: 110-119, Fire et al. (1998) *Nature* 391: 806-811 and Timmons and Fire (1998) *Nature* 395: 854.

Vectors in which RNA encoded by a transcription factor or transcription factor homologue cDNA is over-expressed can also be used to obtain co-suppression of a corresponding endogenous gene, e.g., in the manner described in U.S. Patent No.

5,231,020 to Jorgensen. Such co-suppression (also termed sense suppression) does not require that the entire transcription factor cDNA be introduced into the plant cells, nor does it require that the introduced sequence be exactly identical to the endogenous transcription factor gene of interest. However, as with antisense suppression, the suppressive efficiency will be enhanced as specificity of hybridization is increased, e.g., as the introduced sequence is lengthened, and/or as the sequence similarity between the introduced sequence and the endogenous transcription factor gene is increased.

Vectors expressing an untranslatable form of the transcription factor mRNA, e.g., sequences comprising one or more stop codon, or nonsense mutation) can also be used to suppress expression of an endogenous transcription factor, thereby reducing or eliminating its activity and modifying one or more traits. Methods for producing such constructs are described in U.S. Patent No. 5,583,021. Preferably, such constructs are made by introducing a premature stop codon into the transcription factor gene. Alternatively, a plant trait can be modified by gene silencing using double-strand RNA (Sharp (1999) Genes and Development 13: 139-141). Another method for abolishing the expression of a gene is by insertion mutagenesis using the T-DNA of *Agrobacterium tumefaciens*. After generating the insertion mutants, the mutants can be screened to identify those containing the insertion in a transcription factor or transcription factor homologue gene. Plants containing a single transgene insertion event at the desired gene can be crossed to generate homozygous plants for the mutation. Such methods are well known to those of skill in the art. (See for example Koncz et al. (1992) Methods in Arabidopsis Research, World Scientific.)

Alternatively, a plant phenotype can be altered by eliminating an endogenous gene, such as a transcription factor or transcription factor homologue, e.g., by homologous recombination (Kempin et al. (1997) Nature 389:802-803).

A plant trait can also be modified by using the Cre-lox system (for example, as described in US Pat. No. 5,658,772). A plant genome can be modified to include first and second lox sites that are then contacted with a Cre recombinase. If the lox sites are in the same orientation, the intervening DNA sequence between the two sites

is excised. If the lox sites are in the opposite-orientation, the intervening sequence is inverted.

The polynucleotides and polypeptides of this invention can also be expressed in a plant in the absence of an expression cassette by manipulating the activity or expression level of the endogenous gene by other means. For example, by ectopically expressing a gene by T-DNA activation tagging (Ichikawa et al. (1997) Nature 390 698-701; Kakimoto et al. (1996) Science 274: 982-985). This method entails transforming a plant with a gene tag containing multiple transcriptional enhancers and once the tag has inserted into the genome, expression of a flanking gene coding sequence becomes deregulated. In another example, the transcriptional machinery in a plant can be modified so as to increase transcription levels of a polynucleotide of the invention (See, e.g., PCT Publications WO 96/06166 and WO 98/53057 which describe the modification of the DNA-binding specificity of zinc finger proteins by changing particular amino acids in the DNA-binding motif).

The transgenic plant can also include the machinery necessary for expressing or altering the activity of a polypeptide encoded by an endogenous gene, for example by altering the phosphorylation state of the polypeptide to maintain it in an activated state.

Transgenic plants (or plant cells, or plant explants, or plant tissues) incorporating the polynucleotides of the invention and/or expressing the polypeptides of the invention can be produced by a variety of well established techniques as described above. Following construction of a vector, most typically an expression cassette, including a polynucleotide, e.g., encoding a transcription factor or transcription factor homologue, of the invention, standard techniques can be used to introduce the polynucleotide into a plant, a plant cell, a plant explant or a plant tissue of interest. Optionally, the plant cell, explant or tissue can be regenerated to produce a transgenic plant.

The plant can be any higher plant, including gymnosperms, monocotyledonous and dicotyledonous plants. Suitable protocols are available for *Leguminosae* (alfalfa, soybean, clover, etc.), *Umbelliferae* (carrot, celery, parsnip),

Cruciferae (cabbage, radish, rapeseed, broccoli, etc.), *Curcubitaceae* (melons and cucumber), *Gramineae* (wheat, corn, rice, barley, millet, etc.), *Solanaceae* (potato, tomato, tobacco, peppers, etc.), and various other crops. See protocols described in Ammirato et al. (1984) Handbook of Plant Cell Culture –Crop Species, Macmillan Publ. Co. Shimamoto et al. (1989) Nature 338:274-276; Fromm et al. (1990) Bio/Technology 8:833-839; and Vasil et al. (1990) Bio/Technology 8:429-434.

Transformation and regeneration of both monocotyledonous and dicotyledonous plant cells is now routine, and the selection of the most appropriate transformation technique will be determined by the practitioner. The choice of method will vary with the type of plant to be transformed; those skilled in the art will recognize the suitability of particular methods for given plant types. Suitable methods can include, but are not limited to: electroporation of plant protoplasts; liposome-mediated transformation; polyethylene glycol (PEG) mediated transformation; transformation using viruses; micro-injection of plant cells; micro-projectile bombardment of plant cells; vacuum infiltration; and *Agrobacterium tumefaciens* mediated transformation. Transformation means introducing a nucleotide sequence into a plant in a manner to cause stable or transient expression of the sequence.

Successful examples of the modification of plant characteristics by transformation with cloned sequences which serve to illustrate the current knowledge in this field of technology, and which are herein incorporated by reference, include: U.S. Patent Nos. 5,571,706; 5,677,175; 5,510,471; 5,750,386; 5,597,945; 5,589,615; 5,750,871; 5,268,526; 5,780,708; 5,538,880; 5,773,269; 5,736,369 and 5,610,042.

Following transformation, plants are preferably selected using a dominant selectable marker incorporated into the transformation vector. Typically, such a marker will confer antibiotic or herbicide resistance on the transformed plants, and selection of transformants can be accomplished by exposing the plants to appropriate concentrations of the antibiotic or herbicide.

After transformed plants are selected and grown to maturity, those plants showing a modified trait are identified. The modified trait can be any of those traits described above. Additionally, to confirm that the modified trait is due to changes in

expression levels or activity of the polypeptide or polynucleotide of the invention can be determined by analyzing mRNA expression using Northern blots, RT-PCR or microarrays, or protein expression using immunoblots or Western blots or gel shift assays.

XVII. Integrated Systems – Sequence Identity

Additionally, the present invention may be an integrated system, computer or computer readable medium that comprises an instruction set for determining the identity of one or more sequences in a database. In addition, the instruction set can be used to generate or identify sequences that meet any specified criteria. Furthermore, the instruction set may be used to associate or link certain functional benefits, such improved characteristics, with one or more identified sequence.

For example, the instruction set can include, e.g., a sequence comparison or other alignment program, e.g., an available program such as, for example, the Wisconsin Package Version 10.0, such as BLAST, FASTA, PILEUP, FINDPATTERNS or the like (GCG, Madison, WI). Public sequence databases such as GenBank, EMBL, Swiss-Prot and PIR or private sequence databases such as PHYTOSEQ sequence database (Incyte Genomics, Palo Alto, CA) can be searched.

Alignment of sequences for comparison can be conducted by the local homology algorithm of Smith and Waterman (1981) Adv. Appl. Math. 2:482, by the homology alignment algorithm of Needleman and Wunsch (1970) J. Mol. Biol. 48:443-453, by the search for similarity method of Pearson and Lipman (1988) Proc. Natl. Acad. Sci. U.S.A. 85:2444-2448, by computerized implementations of these algorithms. After alignment, sequence comparisons between two (or more) polynucleotides or polypeptides are typically performed by comparing sequences of the two sequences over a comparison window to identify and compare local regions of sequence similarity. The comparison window can be a segment of at least about 20 contiguous positions, usually about 50 to about 200, more usually about 100 to about 150 contiguous positions. A description of the method is provided in Ausubel et al., *supra*.

A variety of methods for determining sequence relationships can be used, including manual alignment and computer assisted sequence alignment and analysis. This later approach is a preferred approach in the present invention, due to the increased throughput afforded by computer assisted methods. As noted above, a variety of computer programs for performing sequence alignment are available, or can be produced by one of skill.

One example algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul et al. J. Mol. Biol. 215:403-410 (1990). Software for performing BLAST analyses is publicly available, e.g., through the National Center for Biotechnology Information (see internet website at ncbi.nlm.nih.gov). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915). Unless otherwise indicated, "sequence identity" here refers to the % sequence identity generated from a tblastx using the NCBI version of

the algorithm at the default settings using gapped alignments with the filter "off" (see, for example, internet website at ncbi.nlm.nih.gov).

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.,* Karlin & Altschul (1993) Proc. Natl. Acad. Sci. USA 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability ($P(N)$), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence (and, therefore, in this context, homologous) if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, or less than about 0.01, and or even less than about 0.001. An additional example of a useful sequence alignment algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments. The program can align, e.g., up to 300 sequences of a maximum length of 5,000 letters.

The integrated system, or computer typically includes a user input interface allowing a user to selectively view one or more sequence records corresponding to the one or more character strings, as well as an instruction set which aligns the one or more character strings with each other or with an additional character string to identify one or more region of sequence similarity. The system may include a link of one or more character strings with a particular phenotype or gene function. Typically, the system includes a user readable output element that displays an alignment produced by the alignment instruction set.

The methods of this invention can be implemented in a localized or distributed computing environment. In a distributed environment, the methods may implemented on a single computer comprising multiple processors or on a multiplicity of computers. The computers can be linked, e.g. through a common bus, but more preferably the computer(s) are nodes on a network. The network can be a generalized or a dedicated local or wide-area network and, in certain preferred embodiments, the computers may be components of an intra-net or an internet.

Thus, the invention provides methods for identifying a sequence similar or homologous to one or more polynucleotides as noted herein, or one or more target polypeptides encoded by the polynucleotides, or otherwise noted herein and may include linking or associating a given plant phenotype or gene function with a sequence. In the methods, a sequence database is provided (locally or across an inter or intra net) and a query is made against the sequence database using the relevant sequences herein and associated plant phenotypes or gene functions.

Any sequence herein can be entered into the database, before or after querying the database. This provides for both expansion of the database and, if done before the querying step, for insertion of control sequences into the database. The control sequences can be detected by the query to ensure the general integrity of both the database and the query. As noted, the query can be performed using a web browser based interface. For example, the database can be a centralized public database such as those noted herein, and the querying can be done from a remote terminal or computer across an internet or intranet.

XVIII. Examples

The following examples are intended to illustrate but not limit the present invention. The complete descriptions of the traits associated with each polynucleotide of the invention is fully disclosed in Table 4 and Table 6.

Example I: Full Length Gene Identification and Cloning

Putative transcription factor sequences (genomic or ESTs) related to known transcription factors were identified in the *Arabidopsis thaliana* GenBank database using the tblastn sequence analysis program using default parameters and a P-value cutoff threshold of -4 or -5 or lower, depending on the length of the query sequence. Putative transcription factor sequence hits were then screened to identify those containing particular sequence strings. If the sequence hits contained such sequence strings, the sequences were confirmed as transcription factors.

Alternatively, *Arabidopsis thaliana* cDNA libraries derived from different tissues or treatments, or genomic libraries were screened to identify novel members of a transcription family using a low stringency hybridization approach. Probes were

synthesized using gene specific primers in a standard PCR reaction (annealing temperature 60° C) and labeled with ^{32}P dCTP using the High Prime DNA Labeling Kit (Boehringer Mannheim). Purified radiolabelled probes were added to filters immersed in Church hybridization medium (0.5 M NaPO_4 pH 7.0, 7% SDS, 1 % w/v bovine serum albumin) and hybridized overnight at 60°C with shaking. Filters were washed two times for 45 to 60 minutes with 1xSSC, 1% SDS at 60° C.

To identify additional sequence 5' or 3' of a partial cDNA sequence in a cDNA library, 5' and 3' rapid amplification of cDNA ends (RACE) was performed using the U.C. Marathon cDNA amplification kit (Clontech, Palo Alto, CA). Generally, the method entailed first isolating poly(A) mRNA, performing first and second strand cDNA synthesis to generate double stranded cDNA, blunting cDNA ends, followed by ligation of the U.C. Marathon Adaptor to the cDNA to form a library of adaptor-ligated ds cDNA.

Gene-specific primers were designed to be used along with adaptor specific primers for both 5' and 3' RACE reactions. Nested primers, rather than single primers, were used to increase PCR specificity. Using 5' and 3' RACE reactions, 5' and 3' RACE fragments were obtained, sequenced and cloned. The process can be repeated until 5' and 3' ends of the full-length gene were identified. Then the full-length cDNA was generated by PCR using primers specific to 5' and 3' ends of the gene by end-to-end PCR.

Example II: Construction of Expression Vectors

The sequence was amplified from a genomic or cDNA library using primers specific to sequences upstream and downstream of the coding region. The expression vector was pMEN20 or pMEN65, which are both derived from pMON316 (Sanders et al, (1987) Nucleic Acids Research 15:1543-1558) and contain the CaMV 35S promoter to express transgenes. To clone the sequence into the vector, both pMEN20 and the amplified DNA fragment were digested separately with SalI and NotI restriction enzymes at 37° C for 2 hours. The digestion products were subject to electrophoresis in a 0.8% agarose gel and visualized by ethidium bromide staining. The DNA fragments containing the sequence and the linearized plasmid were excised and purified by using a Qiaquick gel extraction kit (Qiagen, Valencia CA). The

fragments of interest were ligated at a ratio of 3:1 (vector to insert). Ligation reactions using T4 DNA ligase (New England Biolabs, Beverly MA) were carried out at 16° C for 16 hours. The ligated DNAs were transformed into competent cells of the *E. coli* strain DH5alpha by using the heat shock method. The transformations were plated on LB plates containing 50 mg/l kanamycin (Sigma, St. Louis, MO). Individual colonies were grown overnight in five milliliters of LB broth containing 50 mg/l kanamycin at 37° C. Plasmid DNA was purified by using Qiaquick Mini Prep kits (Qiagen).

Example III: Transformation of *Agrobacterium* with the Expression Vector

After the plasmid vector containing the gene was constructed, the vector was used to transform *Agrobacterium tumefaciens* cells expressing the gene products. The stock of *Agrobacterium tumefaciens* cells for transformation were made as described by Nagel et al. (1990) FEMS Microbiol Letts. 67: 325-328. *Agrobacterium* strain ABI was grown in 250 ml LB medium (Sigma) overnight at 28°C with shaking until an absorbance (A_{600}) of 0.5 – 1.0 was reached. Cells were harvested by centrifugation at 4,000 x g for 15 min at 4° C. Cells were then resuspended in 250 µl chilled buffer (1 mM HEPES, pH adjusted to 7.0 with KOH). Cells were centrifuged again as described above and resuspended in 125 µl chilled buffer. Cells were then centrifuged and resuspended two more times in the same HEPES buffer as described above at a volume of 100 µl and 750 µl, respectively. Resuspended cells were then distributed into 40 µl aliquots, quickly frozen in liquid nitrogen, and stored at -80° C.

Agrobacterium cells were transformed with plasmids prepared as described above following the protocol described by Nagel et al. For each DNA construct to be transformed, 50 – 100 ng DNA (generally resuspended in 10 mM Tris-HCl, 1 mM EDTA, pH 8.0) was mixed with 40 µl of *Agrobacterium* cells. The DNA/cell mixture was then transferred to a chilled cuvette with a 2mm electrode gap and subject to a 2.5 kV charge dissipated at 25 µF and 200 µF using a Gene Pulser II apparatus (Bio-Rad, Hercules, CA). After electroporation, cells were immediately resuspended in 1.0 ml LB and allowed to recover without antibiotic selection for 2 – 4 hours at 28° C in a shaking incubator. After recovery, cells were plated onto selective medium of LB broth containing 100 µg/ml spectinomycin (Sigma) and incubated for 24-48 hours at 28° C. Single colonies were then picked and inoculated in fresh medium. The

presence of the plasmid construct was verified by PCR amplification and sequence analysis.

Example IV: Transformation of *Arabidopsis* Plants with *Agrobacterium tumefaciens* with Expression Vector

After transformation of *Agrobacterium tumefaciens* with plasmid vectors containing the gene, single *Agrobacterium* colonies were identified, propagated, and used to transform *Arabidopsis* plants. Briefly, 500 ml cultures of LB medium containing 50 mg/l kanamycin were inoculated with the colonies and grown at 28° C with shaking for 2 days until an optical absorbance at 600 nm wavelength over 1 cm (A_{600}) of > 2.0 is reached. Cells were then harvested by centrifugation at 4,000 x g for 10 min, and resuspended in infiltration medium (1/2 X Murashige and Skoog salts (Sigma), 1 X Gamborg's B-5 vitamins (Sigma), 5.0% (w/v) sucrose (Sigma), 0.044 μ M benzylamino purine (Sigma), 200 μ l/l Silwet L-77 (Lehle Seeds) until an A_{600} of 0.8 was reached.

Prior to transformation, *Arabidopsis thaliana* seeds (ecotype Columbia) were sown at a density of ~10 plants per 4" pot onto Pro-Mix BX potting medium (Hummert International) covered with fiberglass mesh (18 mm X 16 mm). Plants were grown under continuous illumination (50-75 μ E/m²/sec) at 22-23° C with 65-70% relative humidity. After about 4 weeks, primary inflorescence stems (bolts) are cut off to encourage growth of multiple secondary bolts. After flowering of the mature secondary bolts, plants were prepared for transformation by removal of all siliques and opened flowers.

The pots were then immersed upside down in the mixture of *Agrobacterium* infiltration medium as described above for 30 sec, and placed on their sides to allow draining into a 1' x 2' flat surface covered with plastic wrap. After 24 h, the plastic wrap was removed and pots are turned upright. The immersion procedure was repeated one week later, for a total of two immersions per pot. Seeds were then collected from each transformation pot and analyzed following the protocol described below.

Example V: Identification of *Arabidopsis* Primary Transformants

Seeds collected from the transformation pots were sterilized essentially as follows. Seeds were dispersed into in a solution containing 0.1% (v/v) Triton X-100 (Sigma) and sterile H₂O and washed by shaking the suspension for 20 min. The wash solution was then drained and replaced with fresh wash solution to wash the seeds for 20 min with shaking. After removal of the second wash solution, a solution containing 0.1% (v/v) Triton X-100 and 70% ethanol (Equistar) was added to the seeds and the suspension was shaken for 5 min. After removal of the ethanol/detergent solution, a solution containing 0.1% (v/v) Triton X-100 and 30% (v/v) bleach (Clorox) was added to the seeds, and the suspension was shaken for 10 min. After removal of the bleach/detergent solution, seeds were then washed five times in sterile distilled H₂O. The seeds were stored in the last wash water at 4° C for 2 days in the dark before being plated onto antibiotic selection medium (1 X Murashige and Skoog salts (pH adjusted to 5.7 with 1M KOH), 1 X Gamborg's B-5 vitamins, 0.9% phytagar (Life Technologies), and 50 mg/l kanamycin). Seeds were germinated under continuous illumination (50-75 $\mu\text{E}/\text{m}^2/\text{sec}$) at 22-23° C. After 7-10 days of growth under these conditions, kanamycin resistant primary transformants (T₁ generation) were visible and obtained. These seedlings were transferred first to fresh selection plates where the seedlings continued to grow for 3-5 more days, and then to soil (Pro-Mix BX potting medium).

Primary transformants were crossed and progeny seeds (T₂) collected; kanamycin resistant seedlings were selected and analyzed. The expression levels of the recombinant polynucleotides in the transformants varies from about a 5% expression level increase to a least a 100% expression level increase. Similar observations are made with respect to polypeptide level expression.

Example VI: Identification of *Arabidopsis* Plants with Transcription Factor Gene Knockouts

The screening of insertion mutagenized *Arabidopsis* collections for null mutants in a known target gene was essentially as described in Krysan et al (1999) Plant Cell 11:2283-2290. Briefly, gene-specific primers, nested by 5-250 base pairs to each other, were designed from the 5' and 3' regions of a known target gene. Similarly, nested sets of primers were also created specific to each of the T-DNA or

transposon ends (the "right" and "left" borders). All possible combinations of gene specific and T-DNA/transposon primers were used to detect by PCR an insertion event within or close to the target gene. The amplified DNA fragments were then sequenced which allows the precise determination of the T-DNA/transposon insertion point relative to the target gene. Insertion events within the coding or intervening sequence of the genes were deconvoluted from a pool comprising a plurality of insertion events to a single unique mutant plant for functional characterization. The method is described in more detail in Yu and Adam, US Application Serial No. 09/177,733 filed October 23, 1998.

Example VII: Identification of Modified Phenotypes in Overexpression or Gene Knockout Plants

Experiments were performed to identify those transformants or knockouts that exhibited modified biochemical characteristics. Among the biochemicals that were assayed were insoluble sugars, such as arabinose, fucose, galactose, mannose, rhamnose or xylose or the like; prenyl lipids, such as lutein, beta-carotene, xanthophyll-1, xanthophyll-2, chlorophylls A or B, or alpha-, delta- or gamma-tocopherol or the like; fatty acids, such as 16:0 (palmitic acid), 16:1 (palmitoleic acid), 18:0 (stearic acid), 18:1 (oleic acid), 18:2 (linoleic acid), 20:0, 18:3 (linolenic acid), 20:1 (eicosenoic acid), 20:2, 22:1 (erucic acid) or the like; waxes, such as by altering the levels of C29, C31, or C33 alkanes; sterols, such as brassicasterol, campesterol, stigmasterol, sitosterol or stigmastanol or the like, glucosinolates, protein or oil levels.

Fatty acids were measured using two methods depending on whether the tissue was from leaves or seeds. For leaves, lipids were extracted and esterified with hot methanolic H_2SO_4 and partitioned into hexane from methanolic brine. For seed fatty acids, seeds were pulverized and extracted in methanol:heptane:toluene:2,2-dimethoxypropane: H_2SO_4 (39:34:20:5:2) for 90 minutes at 80°C. After cooling to room temperature the upper phase, containing the seed fatty acid esters, was subjected to GC analysis. Fatty acid esters from both seed and leaf tissues were analyzed with a Supelco SP-2330 column.

Glucosinolates were purified from seeds or leaves by first heating the tissue at 95°C for 10 minutes. Preheated ethanol:water (50:50) is and after heating at 95°C for a further 10 minutes, the extraction solvent is applied to a DEAE Sephadex column which had been previously equilibrated with 0.5 M pyridine acetate. Desulfoglucosinolates were eluted with 300 ul water and analyzed by reverse phase HPLC monitoring at 226 nm.

For wax alkanes, samples were extracted using an identical method as fatty acids and extracts were analyzed on a HP 5890 GC coupled with a 5973 MSD. Samples were chromatographically isolated on a J&W DB35 mass spectrometer (J&W Scientific).

To measure prenyl lipids levels, seeds or leaves were pulverized with 1 to 2% pyrogallol as an antioxidant. For seeds, extracted samples were filtered and a portion removed for tocopherol and carotenoid/chlorophyll analysis by HPLC. The remaining material was saponified for sterol determination. For leaves, an aliquot was removed and diluted with methanol and chlorophyll A, chlorophyll B, and total carotenoids measured by spectrophotometry by determining optical absorbance at 665.2 nm, 652.5 nm, and 470 nm. An aliquot was removed for tocopherol and carotenoid/chlorophyll composition by HPLC using a Waters uBondapak C18 column (4.6 mm x 150 mm). The remaining methanolic solution was saponified with 10% KOH at 80°C for one hour. The samples were cooled and diluted with a mixture of methanol and water. A solution of 2% methylene chloride in hexane was mixed in and the samples were centrifuged. The aqueous methanol phase was again re-extracted 2% methylene chloride in hexane and, after centrifugation, the two upper phases were combined and evaporated. 2% methylene chloride in hexane was added to the tubes and the samples were then extracted with one ml of water. The upper phase was removed, dried, and resuspended in 400 ul of 2% methylene chloride in hexane and analyzed by gas chromatography using a 50 m DB-5ms (0.25 mm ID, 0.25 um phase, J&W Scientific).

Insoluble sugar levels were measured by the method essentially described by Reiter et al., (1997) Plant Journal 12:335-345. This method analyzes the neutral sugar composition of cell wall polymers found in *Arabidopsis* leaves. Soluble sugars were

separated from sugar polymers by extracting leaves with hot 70% ethanol. The remaining residue containing the insoluble polysaccharides was then acid hydrolyzed with allose added as an internal standard. Sugar monomers generated by the hydrolysis were then reduced to the corresponding alditols by treatment with NaBH₄, then were acetylated to generate the volatile alditol acetates which were then analyzed by GC-FID. Identity of the peaks was determined by comparing the retention times of known sugars converted to the corresponding alditol acetates with the retention times of peaks from wild-type plant extracts. Alditol acetates were analyzed on a Supelco SP-2330 capillary column (30 m x 250 μ m x 0.2 μ m) using a temperature program beginning at 180° C for 2 minutes followed by an increase to 220° C in 4 minutes. After holding at 220° C for 10 minutes, the oven temperature is increased to 240° C in 2 minutes and held at this temperature for 10 minutes and brought back to room temperature.

To identify plants with alterations in total seed oil or protein content, 150mg of seeds from T2 progeny plants were subjected to analysis by Near Infrared Reflectance Spectroscopy (NIRS) using a Foss NirSystems Model 6500 with a spinning cup transport system. NIRS is a non-destructive analytical method used to determine seed oil and protein composition. Infrared is the region of the electromagnetic spectrum located after the visible region in the direction of longer wavelengths. 'Near infrared' owns its name for being the infrared region near to the visible region of the electromagnetic spectrum. For practical purposes, near infrared comprises wavelengths between 800 and 2500 nm. NIRS is applied to organic compounds rich in O-H bonds (such as moisture, carbohydrates, and fats), C-H bonds (such as organic compounds and petroleum derivatives), and N-H bonds (such as proteins and amino acids). The NIRS analytical instruments operate by statistically correlating NIRS signals at several wavelengths with the characteristic or property intended to be measured. All biological substances contain thousands of C-H, O-H, and N-H bonds. Therefore, the exposure to near infrared radiation of a biological sample, such as a seed, results in a complex spectrum which contains qualitative and quantitative information about the physical and chemical composition of that sample.

The numerical value of a specific analyte in the sample, such as protein content or oil content, is mediated by a calibration approach known as chemometrics.

Chemometrics applies statistical methods such as multiple linear regression (MLR), partial least squares (PLS), and principle component analysis (PCA) to the spectral data and correlates them with a physical property or other factor, that property or factor is directly determined rather than the analyte concentration itself. The method first provides "wet chemistry" data of the samples required to develop the calibration.

Calibration for Arabidopsis seed oil composition was performed using accelerated solvent extraction using 1 g seed sample size and was validated against certified canola seed. A similar wet chemistry approach was performed for seed protein composition calibration.

Data obtained from NIRS analysis was analyzed statistically using a nearest-neighbor (N-N) analysis. The N-N analysis allows removal of within-block spatial variability in a fairly flexible fashion which does not require prior knowledge of the pattern of variability in the chamber. Ideally, all hybrids are grown under identical experimental conditions within a block (rep). In reality, even in many block designs, significant within-block variability exists. Nearest-neighbor procedures are based on assumption that environmental effect of a plot is closely related to that of its neighbors. Nearest-neighbor methods use information from adjacent plots to adjust for within-block heterogeneity and so provide more precise estimates of treatment means and differences. If there is within-plot heterogeneity on a spatial scale that is larger than a single plot and smaller than the entire block, then yields from adjacent plots will be positively correlated. Information from neighboring plots can be used to reduce or remove the unwanted effect of the spatial heterogeneity, and hence improve the estimate of the treatment effect. Data from neighboring plots can also be used to reduce the influence of competition between adjacent plots. The Papadakis N-N analysis can be used with designs to remove within-block variability that would not be removed with the standard split plot analysis (Papadakis, 1973, Inst. d'Amelior. Plantes Thessaloniki (Greece) Bull. Scientif., No. 23; Papadakis, 1984, Proc. Acad. Athens, 59, 326-342).

Experiments were performed to identify those transformants or knockouts that exhibited an improved pathogen tolerance. For such studies, the transformants were exposed to biotrophic fungal pathogens, such as *Erysiphe orontii*, and necrotrophic

fungus pathogens, such as *Fusarium oxysporum*. *Fusarium oxysporum* isolates cause vascular wilts and damping off of various annual vegetables, perennials and weeds (Mauch-Mani and Slusarenko (1994) Molecular Plant-Microbe Interactions 7: 378-383). For *Fusarium oxysporum* experiments, plants grown on Petri dishes were sprayed with a fresh spore suspension of *F. oxysporum*. The spore suspension was prepared as follows: A plug of fungal hyphae from a plate culture was placed on a fresh potato dextrose agar plate and allowed to spread for one week. 5 ml sterile water was then added to the plate, swirled, and pipetted into 50 ml Armstrong *Fusarium* medium. Spores were grown overnight in *Fusarium* medium and then sprayed onto plants using a Preval paint sprayer. Plant tissue was harvested and frozen in liquid nitrogen 48 hours post infection.

Erysiphe orontii is a causal agent of powdery mildew. For *Erysiphe orontii* experiments, plants were grown approximately 4 weeks in a greenhouse under 12 hour light (20°C, ~30% relative humidity (rh)). Individual leaves were infected with *E. orontii* spores from infected plants using a camel's hair brush, and the plants were transferred to a Percival growth chamber (20°C, 80% rh.). Plant tissue was harvested and frozen in liquid nitrogen 7 days post infection.

Botrytis cinerea is a necrotrophic pathogen. *Botrytis cinerea* was grown on potato dextrose agar in the light. A spore culture was made by spreading 10 ml of sterile water on the fungus plate, swirling and transferring spores to 10 ml of sterile water. The spore inoculum (approx. 105 spores/ml) was used to spray 10 day-old seedlings grown under sterile conditions on MS (minus sucrose) media. Symptoms were evaluated every day up to approximately 1 week.

Infection with bacterial pathogens *Pseudomonas syringae* pv *maculicola* (Psm) strain 4326 and pv *maculicola* strain 4326 was performed by hand inoculation at two doses. Two inoculation doses allows the differentiation between plants with enhanced susceptibility and plants with enhanced resistance to the pathogen. Plants were grown for 3 weeks in the greenhouse, then transferred to the growth chamber for the remainder of their growth. Psm ES4326 was hand inoculated with 1 ml syringe on 3 fully-expanded leaves per plant (4 1/2 wk old), using at least 9 plants per overexpressing line at two inoculation doses, OD=0.005 and OD=0.0005. Disease

scoring occurred at day 3 post-inoculation with pictures of the plants and leaves taken in parallel.

In some instances, expression patterns of the pathogen-induced genes (such as defense genes) was monitored by microarray experiments. cDNAs were generated by PCR and resuspended at a final concentration of ~ 100 ng/ul in 3X SSC or 150mM Na-phosphate (Eisen and Brown (1999) *Methods Enzymol.* 303:179-205). The cDNAs were spotted on microscope glass slides coated with polylysine. The prepared cDNAs were aliquoted into 384 well plates and spotted on the slides using an x-y-z gantry (OmniGrid) purchased from GeneMachines (Menlo Park, CA) outfitted with quill type pins purchased from Telechem International (Sunnyvale, CA). After spotting, the arrays were cured for a minimum of one week at room temperature, rehydrated and blocked following the protocol recommended by Eisen and Brown (1999; *supra*).

Sample total RNA (10 ug) samples were labeled using fluorescent Cy3 and Cy5 dyes. Labeled samples were resuspended in 4X SSC/0.03% SDS/4 ug salmon sperm DNA/2 ug tRNA/ 50mM Na-pyrophosphate, heated for 95°C for 2.5 minutes, spun down and placed on the array. The array was then covered with a glass coverslip and placed in a sealed chamber. The chamber was then kept in a water bath at 62°C overnight. The arrays were washed as described in Eisen and Brown (1999) and scanned on a General Scanning 3000 laser scanner. The resulting files are subsequently quantified using Imagene, a software purchased from BioDiscovery (Los Angeles, CA).

Experiments were performed to identify those transformants or knockouts that exhibited an improved environmental stress tolerance. For such studies, the transformants were exposed to a variety of environmental stresses. Plants were exposed to chilling stress (6 hour exposure to 4-8° C), heat stress (6 hour exposure to 32-37° C), high salt stress (6 hour exposure to 200 mM NaCl), drought stress (168 hours after removing water from trays), osmotic stress (6 hour exposure to 3 M mannitol), or nutrient limitation (nitrogen, phosphate, and potassium) (Nitrogen: all components of MS medium remained constant except N was reduced to 20 mg/l of NH₄NO₃, or Phosphate: All components of MS medium except KH₂PO₄, which was

replaced by K_2SO_4 , Potassium: All components of MS medium except removal of KNO_3 and KH_2PO_4 , which were replaced by NaH_4PO_4).

Experiments were performed to identify those transformants or knockouts that exhibited a modified structure and development characteristics. For such studies, the transformants were observed by eye to identify novel structural or developmental characteristics associated with the ectopic expression of the polynucleotides or polypeptides of the invention.

Experiments were performed to identify those transformants or knockouts that exhibited modified sugar-sensing. For such studies, seeds from transformants were germinated on media containing 5% glucose or 9.4% sucrose which normally partially restrict hypocotyl elongation. Plants with altered sugar sensing may have either longer or shorter hypocotyls than normal plants when grown on this media. Additionally, other plant traits may be varied such as root mass.

Flowering time was measured by the number of rosette leaves present when a visible inflorescence of approximately 3 cm is apparent. Rosette and total leaf number on the progeny stem are tightly correlated with the timing of flowering (Koornneef et al (1991) *Mol. Gen. Genet* 229:57-66). The vernalization response was measured. For vernalization treatments, seeds were sown to MS agar plates, sealed with micropore tape, and placed in a 4°C cold room with low light levels for 6-8 weeks. The plates were then transferred to the growth rooms alongside plates containing freshly sown non-vernalized controls. Rosette leaves were counted when a visible inflorescence of approximately 3 cm was apparent.

Modified phenotypes observed for particular overexpressor or knockout plants are provided in Table 4. For a particular overexpressor that shows a less beneficial characteristic, it may be more useful to select a plant with a decreased expression of the particular transcription factor. For a particular knockout that shows a less beneficial characteristic, it may be more useful to select a plant with an increased expression of the particular transcription factor.

The sequences of the Sequence Listing or those in Tables 4 , 5 or those disclosed here can be used to prepare transgenic plants and plants with altered traits. The specific transgenic plants listed below are produced from the sequences of the Sequence Listing, as noted. Table 4 provides exemplary polynucleotide and polypeptide sequences of the invention. Table 4 includes, from left to right for each sequence: the first column shows the polynucleotide SEQ ID NO; the second column shows the Mendel Gene ID No., GID; the third column shows the trait(s) resulting from the knock out or overexpression of the polynucleotide in the transgenic plant; the fourth column shows the category of the trait; the fifth column shows the transcription factor family to which the polynucleotide belongs; the sixth column ("Comment"), includes specific effects and utilities conferred by the polynucleotide of the first column; the seventh column shows the SEQ ID NO of the polypeptide encoded by the polynucleotide; and the eighth column shows the amino acid residue positions of the conserved domain in amino acid (AA) co-ordinates.

Seed of plants overexpressing sequences G265 (SEQ ID NOs:871 and 872), G715 (SEQ ID NOs:925 and 926), G1471 (SEQ ID NOs:311 and 312), G1793 (SEQ ID NOs:365 and 366), G1838 (SEQ ID NOs:381 and 382), G1902 (SEQ ID NOs:405 and 406), G286 (SEQ ID NOs:877 and 878), G2138 (SEQ ID NOs:865 and 866) and G2830 (SEQ ID NOs:875 and 876) was subjected to NIR analysis and a significant increase in seed oil content compared with seed from control plants was identified.

G192: G192 (SEQ ID NO: 859) was expressed in all plant tissues and under all conditions examined. Its expression was slightly induced upon infection by *Fusarium*. G192 was analyzed using transgenic plants in which this gene was expressed under the control of the 35S promoter. G192 overexpressors were late flowering under 12 hour light and had more leaves than control plants. This phenotype was manifested in the three T2 lines analyzed. Results of one experiment suggest that G192 overexpressor was more susceptible to infection with a moderate dose of the fungal pathogen *Erysiphe orontii*. The decrease in seed oil observed for one line was replicated in an independent experiment. G192 overexpression delayed flowering. A wide variety of applications exist for systems that either lengthen or shorten the time to flowering, or for systems of inducible flowering time control. In particular, in species where the vegetative parts of the plants constitute the crop and

the reproductive tissues are discarded, it will be advantageous to delay or prevent flowering. Extending vegetative development can bring about large increases in yields. G192 can be used to manipulate the defense response in order to generate pathogen-resistant plants. G192 can be used to manipulate seed oil content, which can be of nutritional value.

Closely Related Genes from Other Species

G192 had some similarity within the conserved WRKY domain to non-Arabidopsis plant proteins.

G1946: G1946 (SEQ ID NO: 801) was studied using transgenic plants in which the gene was expressed under the control of the 35S promoter. Overexpression of G1946 resulted in accelerated flowering; with 35S::G1946 transformants producing flower buds up to a week earlier than wild-type controls (24-hour light conditions). These effects were seen in 12/20 primary transformants and in two independent plantings of each of the three T2 lines. Unlike many early flowering Arabidopsis transgenic lines, which are dwarfed, 35S::G1946 transformants often reached full-size at maturity, and produced large quantities of seeds, although the plants were slightly pale in coloration and had slightly flat leaves compared to wild-type. In addition, 35S::G1946 plants showed an altered response to phosphate deprivation. Seedlings of G1946 overexpressor plants showed more secondary root growth on phosphate-free media, when compared to wild-type control. In a repeat experiment, all three lines showed the phenotype. Overexpression of G1946 in Arabidopsis also resulted in an increase in seed glucosinolate M39501 in T2 lines 1 and 3. An increase in seed oil and a decrease in seed protein was also observed in these two lines. G1946 was ubiquitously expressed, and does not appear to be significantly induced or repressed by any of the biotic and abiotic stress conditions tested at this time, with the exception of cold, which repressed G1946 expression. G1946 can be used to modify flowering time, as well as to improve the plant's performance in conditions of limited phosphate, and to alter seed oil, protein, and glucosinolate composition.

Closely Related Genes from Other Species

A comparison of the amino acid sequence of G1946 with sequences available from GenBank showed strong similarity with plant HSFs of several species (*Lycopersicon peruvianum*, *Medicago truncatula*, *Lycopersicon esculentum*, *Glycine max*, *Solanum tuberosum*, *Oryza sativa* and *Hordeum vulgare* subsp. *vulgare*).

G375: The sequence of G375 (SEQ ID NO:239) was experimentally determined and G375 was analyzed using transgenic plants in which G375 was expressed under the control of the 35S promoter. Overexpression of G375 produced marked effects on leaf development. At early stages of growth, 35S::G375 seedlings developed narrow, upward pointing leaves with long petioles (possibly indicating a disruption in circadian-clock controlled processes or nyctinastic movements). Additionally, some seedlings were noted to have elongated hypocotyls, and some were rather small compared to wild-type controls. Comparable phenotypes were obtained by overexpression of an AP2 family gene, G2113 (SEQ ID NO: 85). Following the switch to flowering, 35S::G375 plants showed reduced fertility, which possibly arose from a failure of stamens to fully elongate. One of the three T2 lines, (#41) was later flowering than wild-type controls, and also developed large numbers of small secondary rosette leaves in the axils of the primary rosette. Although these effects were not noted in the other two lines, the phenotypes obtained in line 41 were somewhat similar to those produced by overexpression of another Z-dof gene, G736 (SEQ ID NO: 211). G375 was expressed in all tissues, although at different levels. It was expressed at low levels in the root and germinating seed, and expressed at high levels in the embryo. The effects of G375 on leaf architecture are of potential interest to the ornamental horticulture industry.

Closely Related Genes from Other Species

G375 showed some homology to non-Arabidopsis plant proteins within the conserved Dof domain.

G1255: The sequence of G1255 (SEQ ID NO: 273) was experimentally determined and G1255 was analyzed using transgenic plants in which G1255 was expressed under the control of the 35S promoter. Plants overexpressing G1255 had

alterations in leaf architecture, a reduction in apical dominance, an increase in seed size, and showed more disease symptoms following inoculation with a low dose of the fungal pathogen *Botrytis cinerea*. G1255 was constitutively expressed and not significantly induced by any conditions tested. On the basis of the phenotypes produced by overexpression of G1255, G1255 can be used to manipulate the plant's defense response to produce pathogen resistance, alter plant architecture, or alter seed size.

Closely Related Genes from Other Species

G1255 showed strong homology to a putative rice zing finger protein represented by sequence AC087181_3. Sequence identity between these two protein extended beyond the conserved domain, and therefore, these genes can be orthologs.

G865: The complete cDNA sequence of G865 (SEQ ID NO: 557) was determined. G865 was ubiquitously expressed in *Arabidopsis* tissues. G865 was analyzed using transgenic plants in which G865 was expressed under the control of the 35S promoter. Plants overexpressing G865 were early flowering, with numerous secondary inflorescence meristems giving them a bushy appearance. G865 overexpressors were more susceptible to infection with a moderate dose of the fungal pathogens *Erysiphe orontii* and *Botrytis cinerea*. In addition, seeds from G865 overexpressing plants showed a trend of increased protein and reduced oil content, although the observed changes were not beyond the criteria used for judging significance except in one line. G865 can be used to control flowering time. G865 can be used to manipulate the defense response in order to generate pathogen-resistant plants. G865 can be used to alter seed oil and protein content of a plant.

Closely Related Genes from Other Species

G865 and other non-*Arabidopsis* AP2/EREBP proteins were similar within the conserved AP2 domain.

G2509: G2509 (SEQ ID NO: 23) was studied using transgenic plants in which the gene was expressed under the control of the 35S promoter. Overexpression of G2509 caused multiple alterations in plant growth and development, most notably, altered branching patterns, and a reduction in apical dominance, giving the plants a

shorter, more bushy stature than wild type. Twenty 35S::G2509 primary transformants were examined; at early stages of rosette development, these plants displayed a wild-type phenotype. However, at the switch to flowering, almost all T1 lines showed a marked loss of apical dominance and large numbers of secondary shoots developed from axils of primary rosette leaves. In the most extreme cases, the shoots had very short internodes, giving the inflorescence a very bushy appearance. Such shoots were often very thin and flowers were relatively small and poorly fertile. At later stages, many plants appeared very small and had a low seed yield compared to wild type. In addition to the effects on branching, a substantial number of 35S::G2509 primary transformants also flowered early and had buds visible several days prior to wild type. Similar effects on inflorescence development were noted in each of three T2 populations examined. The branching and plant architecture phenotypes observed in 35S::G2509 lines resemble phenotypes observed for three other AP2/EREBP genes: G865 (SEQ ID NO: 557), G1411 (SEQ ID NO: 3), and G1794 (SEQ ID NO: 13). G2509, G865, and G1411 form a small clade within the large AP2/EREBP family, and G1794, although not belonging to the clade, is one of the AP2/EREBP genes closest to it in the phylogenetic tree. It is thus likely that all these genes share a related function, such as affecting hormone balance. Overexpression of G2509 in Arabidopsis resulted in an increase in alpha-tocopherol in seeds in T2 lines 5 and 11. G2509 was ubiquitously expressed in Arabidopsis plant tissue. G2509 expression levels were altered by a variety of environmental or physiological conditions. G2509 can be used to manipulate plant architecture and development. G2509 can be used to alter tocopherol composition. Tocopherols have anti-oxidant and vitamin E activity. G2509 can be useful in altering flowering time. A wide variety of applications exist for systems that either lengthen or shorten the time to flowering.

Closely Related Genes from Other Species

G2509 showed some sequence similarity with known genes from other plant species within the conserved AP2/EREBP domain.

G2347: G2347 (SEQ ID NO: 1119) was analyzed using transgenic plants in which G2347 was expressed under the control of the 35S promoter. Overexpression of G2347 markedly reduced the time to flowering in Arabidopsis. This phenotype

was apparent in the majority of primary transformants and in all plants from two out of the three T2 lines examined. Under continuous light conditions, 35S::G2347 plants formed flower buds up a week earlier than wild type. Many of the plants were rather small and spindly compared to controls. To demonstrate that overexpression of G2347 could induce flowering under less inductive photoperiods, two T2 lines were re-grown in 12 hour conditions; again, all plants from both lines bolted early, with some initiating flower buds up to two weeks sooner than wild-type. As determined by RT-PCR, G2347 was highly expressed in rosette leaves and flowers, and to much lower levels in embryos and siliques. No expression of G2347 was detected in the other tissues tested. G2347 expression was repressed by cold, and by auxin treatments and by infection by Erysiphe. G2347 is also highly similar to the Arabidopsis protein G2010 (SEQ ID NO: 1121). The level of homology between these two proteins suggested they could have similar, overlapping, or redundant functions in Arabidopsis. In support of this hypothesis, overexpression of both G2010 and G2347 resulted in early flowering phenotypes in transgenic plants.

Closely Related Genes from Other Species

The closest relative to G2347 is the Antirrhinum protein, SBP2 (CAA63061). The similarity between these two proteins is extensive enough to suggest they might have similar functions in a plant.

G988: G988 (SEQ ID NO: 43) was analyzed using transgenic plants in which G988 was expressed under the control of the 35S promoter. Plants overexpressing G988 had multiple morphological phenotypes. The transgenic plants were generally smaller than wild-type plants, had altered leaf, inflorescence and flower development, altered plant architecture, and altered vasculature. In one transgenic line overexpressing G988 (line 23), an increase in the seed glucosinolate M39489 was observed. The phenotype of plants overexpressing G988 was wild-type in all other assays performed. In wild-type plants, G988 was expressed primarily in flower and silique tissue, but was also present at detectable levels in all other tissues tested. Expression of G988 was induced in response to heat treatment, and repressed in response to infection with Erysiphe. Based on the observed morphological phenotypes of the transgenic plants, G988 can be used to create plants with larger flowers. This can have value in the ornamental horticulture industry. The reduction

in the formation of lateral branches suggests that G988 can have utility on the forestry industry. The Arabidopsis plants overexpressing G988 also had reduced fertility. This can be a desirable trait in some instances, as it can be exploited to prevent or minimize the escape of GMO (genetically modified organism) pollen into the environment.

Closely Related Genes from Other Species

The amino acid sequence for the *Capsella rubella* hypothetical protein represented by GenBank accession number CRU303349 was significantly identical to G988 outside of the SCR conserved domains. The *Capsella rubella* hypothetical protein is 90% identical to G988 over a stretch of roughly 450 amino acids. Therefore, it is likely that the *Capsella rubella* gene is an ortholog of G988.

G2346: G2346 (SEQ ID NO: 459) was analyzed using transgenic plants in which the gene was expressed under the control of the 35S promoter. 35S::G2346 seedlings from all three T2 populations had slightly larger cotyledons and appeared somewhat more advanced than controls. This indicated that the seedlings developed more rapidly than the control plants. At later stages, however, G2346 overexpressing plants showed no consistent differences from control plants. The phenotype of these transgenic plants was wild-type in all other assays performed. According to RT-PCR analysis, G2346 is expressed ubiquitously.

Closely Related Genes from Other Species

G2346 shows some sequence similarity with known genes from other plant species within the conserved SBP domain.

G1354: The complete sequence of G1354 (SEQ ID NO: 285) was determined. G1354 was analyzed using transgenic plants in which G1354 was expressed under the control of the 35S promoter. Overexpression of G1354 produced highly deleterious effects on growth and development. Only three 35S::G1354 T1 plants were obtained; all were extremely tiny and slow developing. After three weeks of growth, each of the plants comprised a completely disorganized mass of leaves and root that had no clear axis of growth. Since these individuals would not have survived transplantation to soil, they were harvested for RT-PCR analysis; all three plants showed moderate

levels of G1354 overexpression compared to whole wild-type seedlings of an equivalent size. Only a very small number of transformants were obtained from two selection attempts on separate batches of T0 seed. Usually between 15 and 120 transformants are obtained from each aliquot of 300 mg T0 seed from wild-type plants. The low transformation frequency obtained in this experiment suggests that high levels of G1354 overexpression might have completely lethal effects and prevent transformed seeds from germinating. As determined by RT-PCR, G1354 was uniformly expressed in all tissues and under all conditions tested in RT-PCR. However, the gene was repressed in leaf tissue in response to Erysiphe infection.

Closely Related Genes from Other Species

G1354 is closely related to a NAM protein encoded by polynucleotide from rice (AC005310). Similarity between G1354 and this rice protein extends beyond the signature motif of the family to a level that would suggest the genes are orthologs.

G1063: G1063 (SEQ ID NO: 119) is a member of a clade of highly related HLH/MYC proteins that also includes G779 (SEQ ID NO: 113), G1499 (SEQ ID NO: 7), G2143 (SEQ ID NO: 129), and G2557 (SEQ ID NO: 133). All of these genes caused similar pleiotropic phenotypic effects when overexpressed, the most striking of which was the production of ectopic carpelloid tissue. These genes can be considered key regulators of carpel development. A spectrum of developmental alterations was observed amongst 35S::G1063 primary transformants and the majority were markedly small, dark green, and had narrow curled leaves. The most severely affected individuals were completely sterile and formed highly abnormal inflorescences; shoots often terminated in pin-like structures, and flowers were replaced by filamentous carpelloid structures. In other cases, flowers showed internode elongation between floral whorls, with a central carpel protruding on a pedicel-like organ. Additionally, lateral branches sometimes failed to develop and tiny patches of carpelloid tissue formed at axillary nodes of the inflorescence. In lines with an intermediate phenotype, flowers contained defined whorls of organs, but sepals were converted to carpelloid structures or displayed patches of carpelloid tissue. In contrast, lines with a weak phenotype developed relatively normal flowers and produced a reasonable quantity of seed. Such plants were still distinctly smaller than wild-type controls. Since the strongest 35S::G1063 lines were sterile, three lines

with a relatively weak phenotype, that had produced sufficient seed for biochemical and physiological analysis, were selected for further study. Two of the T2 populations (T2-28,37) were clearly small, darker green and possessed narrow leaves compared to wild type. Plants from one of these populations (T2-28) also produced occasional branches with abnormal flowers like those seen in the T1. The final T2 population (T2-30) displayed a very mild phenotype. Overexpression of G1063 in *Arabidopsis* resulted in a decrease in seed oil content in T2 lines 28 and 37. No altered phenotypes were detected in any of the physiological assays, except that the plants were noted to be somewhat small and produce anthocyanin when grown in Petri plates. G1063 was expressed at low to moderate levels in roots, flowers, rosette leaves, embryos, and germinating seeds, but was not detected in shoots or siliques. It was induced by auxin. G1063 can be used to manipulate flower form and structure or plant fertility. One application for manipulation of flower structure can be in the production of saffron, which is derived from the stigmas of *Crocus sativus*. G1063 has utility in manipulating seed oil and protein content.

Closely Related Genes from Other Species

G1063 protein shared extensive homology in the basic helix loop helix region with a protein sequence encoded by Glycine max cDNA clone (AW832545) as well as a tomato root, plants pre-anthesis *Lycopersicon esculentum* cDNA (BE451174).

G2143: G2143 (SEQ ID NO: 129) is a member of a clade of highly related HLH/MYC proteins that also includes G779 (SEQ ID NO: 113), G1063 (SEQ ID NO: 119), G1499 (SEQ ID NO: 7), and G2557 (SEQ ID NO: 133). All of these genes caused similar pleiotropic phenotypic effects when overexpressed, the most striking of which was the production of ectopic carpelloid tissue. These genes can be considered key regulators of carpel development. Twelve out of twenty 35S::G2143 T1 lines showed a very severe phenotype; these plants were markedly small and had narrow, curled, dark-green leaves. Such individuals were completely sterile and formed highly abnormal inflorescences; shoots often terminated in pin-like structures, and flowers were replaced by filamentous carpelloid structures, or a fused mass of carpelloid tissue. Furthermore, lateral branches usually failed to develop, and tiny patches of stigmatic tissue often formed at axillary nodes of the inflorescence. Strongly affected plants displayed the highest levels of transgene expression

(determined by RT-PCR). The remaining T1 lines showed lower levels of G2143 overexpression; these plants were still distinctly smaller than wild type, but had relatively normal inflorescences and produced seed. Since the strongest 35S::G2143 lines were sterile, three lines with a relatively weak phenotype, that had produced sufficient seed for biochemical analysis, were selected for further study. T2-11 plants displayed a very mild phenotype and had somewhat small, narrow, dark green leaves. The other two T2 populations, however, appeared wild-type, suggesting that transgene activity might have been reduced between the generations. Reduced seedling vigor was noted in the physiological assays. G2143 expression was detected at low levels in flowers and siliques, and at higher levels in germinating seed. G2143 can be used to manipulate flower form and structure or plant fertility. One application for manipulation of flower structure can be in the production of saffron, which is derived from the stigmas of *Crocus sativus*.

Closely Related Genes from Other Species

G2143 protein shared extensive homology in the basic helix loop helix region with a protein encoded by Glycine max cDNA clones (AW832545, BG726819 and BG154493) and a *Lycopersicon esculentum* cDNA clone (BE451174). There was lower homology outside of the region.

G2557: G2557 (SEQ ID NO: 133) is a member of a clade of highly related HLH/MYC proteins that also includes G779 (SEQ ID NO: 113), G1063 (SEQ ID NO: 119), G1499 (SEQ ID NO: 7), and G2143 (SEQ ID NO: 129). All of these genes caused similar pleiotropic phenotypic effects when overexpressed, the most striking of which was the production of ectopic carpelloid tissue. These genes can be considered key regulators of carpel development. The flowers of 35S::G2557 primary transformants displayed patches of stigmatic papillae on the sepals, and often had rather narrow petals and poorly developed stamens. Additionally, carpels were also occasionally held outside of the flower at the end of an elongated pedicel like structure. As a result of such defects, 35S::G2557 plants often showed very poor fertility and formed small wrinkled siliques. In addition to such floral abnormalities, the majority of primary transformants were also small and darker green in coloration than wild type. Approximately one third of the T1 plants were extremely tiny and completely sterile. Three T1 lines (#7,9,12), that had produced some seeds, and

showed a relatively weak phenotype, were chosen for further study. All three of the T2 populations from these lines contained plants that were distinctly small, had abnormal flowers, and were poorly fertile compared to controls. Stigmatic tissue was not noted on the sepals of plants from these three T2 lines. Another line (#4) that had shown a moderately strong phenotype in the T1 was sown for only morphological analysis in the T2 generation. T2-4 plants were small, dark green, and produced abnormal flowers with ectopic stigmatic tissue on the sepals, as had been seen in the parental plant. G2557 expression was detected at low to moderate levels in all tissues tested except shoots. It was induced by cold, heat, and salt, and repressed by pathogen infection. G1063 can be used to manipulate flower form and structure or plant fertility. One application for manipulation of flower structure can be in the production of saffron, which is derived from the stigmas of *Crocus sativus*.

Closely Related Genes from Other Species

G2557 protein shows extensive sequence similarity in the region of basic helix loop helix with a protein encoded by Glycine max cDNA clone (BE347811).

G2430: The complete sequence of G2430 (SEQ ID NO: 697) was determined. G2430 is a member of the response regulator class of GARP proteins (ARR genes), although one of the two conserved aspartate residues characteristic of response regulators is not present. The second aspartate, the putative phosphorylated site, is retained so G2430 can have response regulator function. G2430 is specifically expressed in embryo and silique tissue. In morphological analyses, plants overexpressing G2430 showed more rapid growth than control plants at early stages, and in two of three lines examined produced large, flat leaves. Early flowering was observed for some lines, but this effect was inconsistent between plantings. G2430 can regulate plant growth. Overexpression of G2430 in *Arabidopsis* also resulted in seedlings that are slightly more tolerant to heat in a germination assay. Seedlings from G2430 overexpressing transgenic plants were slightly greener than the control seedlings under high temperature conditions. In a repeat experiment on individual lines, G2430 line 15 showed the strongest heat tolerant phenotype. G2430 can be useful to promote faster development and reproduction in plants.

Closely Related Genes from Other Species

G2430 had some similarity within of the conserved GARP and response-regulator domains to non-Arabidopsis proteins.

G1478: The sequence of G1478 (SEQ ID NO: 831) was determined and G1478 was analyzed using transgenic plants in which G1478 was expressed under the control of the 35S promoter. Plants overexpressing G1478 had a general delay in progression through the life cycle, in particular a delay in flowering time. G1478 is expressed at higher levels in flowers, rosettes and embryos but otherwise expression is constitutive. Based on the phenotypes produced through G1478 overexpression, G1478 can be used to manipulate the rate at which plants grow, and flowering time.

Closely Related Genes from Other Species

G1478 shows some homology to non-Arabidopsis proteins within the conserved domain.

G681: G681 (SEQ ID NO: 579) was analyzed using transgenic plants in which the gene was expressed under the control of the 35S promoter. Approximately half of the 35S::G681 primary transformants were markedly small and formed narrow leaves compared to controls. These plants often produced thin inflorescence stems, had rather poorly formed flowers with low pollen production, and set few seeds. Three T1 lines with relatively weak phenotypes, which had produced reasonable quantities of seed, were selected for further study. Plants from one of the T2 populations were noted to be slightly small, but otherwise the T2 lines displayed no consistent differences in morphology from controls. In leaves of two of the T2 lines, overexpression of G681 resulted in an increase in the percentage of the glucosinolate M39480. According to RT-PCR analysis, G681 expression was detected at very low levels in flower and rosette leaf tissues. G681 was induced by drought stress. G681 can be used to alter glucosinolate composition in plants. Increases or decreases in specific glucosinolates or total glucosinolate content are desirable depending upon the particular application. For example: (1) Glucosinolates are undesirable components of the oilseeds used in animal feed, since they produce toxic effects. Low-glucosinolate varieties of canola have been developed to combat this problem. (2) Some glucosinolates have anti-cancer activity; thus, increasing the levels or

composition of these compounds might be of interest from a nutraceutical standpoint. (3) Glucosinolates form part of a plants natural defense against insects. Modification of glucosinolate composition or quantity could therefore afford increased protection from predators. Furthermore, in edible crops, tissue specific promoters can be used to ensure that these compounds accumulate specifically in tissues, such as the epidermis, which are not taken for consumption.

Closely Related Genes from Other Species

G681 shows some sequence similarity with known genes from other plant species within the conserved Myb domain.

G878: G878 (SEQ ID NO: 611) was studied using transgenic plants in which the gene was expressed under the control of the 35S promoter. Analysis of primary transformants revealed that overexpression of G878 delays the onset of flowering in Arabidopsis. 11/20 of the 35S::G878 T1 plants flowered approximately one week later than wild type under continuous light conditions. These plants were also darker green, had shorter stems, and senesced later than controls. G878 was ubiquitously expressed. G878 can be used to modify flowering time and senescence, and a wide variety of applications exist for systems that either lengthen or shorten the time to flowering.

Closely Related Genes from Other Species

G878 was highly related to other WRKY proteins from a variety of plant species, such as the Nicotiana tabacum DNA-binding protein 2 (WRKY2) (AF096299), and a Cucumis sativus SPF1-like DNA-binding protein (L44134).

G374: G374 (SEQ ID NO: 47) was expressed at low levels throughout the plant and was induced by salicylic acid. G374 was investigated using lines carrying a T-DNA insertion in this gene. The T-DNA insertion was approximately three quarters of the way into the protein coding sequence and should result in a null mutation. Homozygosity for a T-DNA insertion within G374 caused lethality at early stages of embryo development. In an initial screen for G374 knockouts, heterozygous plants were identified. Seed from those individuals was sown to soil and eleven plants were PCR-screened to identify homozygotes. No homozygotes were obtained;

6 of the progeny were heterozygous whilst the other 5 were wild type. This raised the prospect that homozygosity for the G374 insertion was lethal. To examine this possibility further, heterozygous KO.G374 plants were re-grown. These individuals looked wild type, but their siliques were examined for seed abnormalities. When green siliques were dissected, around 25% of developing seeds were white or aborted. Embryos from these siliques were cleared using Hoyers solution, and examined under the microscope. It was apparent that embryos from the white seeds had arrested at early (globular or heart) stages of development, whilst embryos from the normal seeds were fully developed. Such arrested or aborted seeds most likely represented homozygotes for the G374 insertion. To support this conclusion, seed was collected from heterozygous plants and sown to kanamycin plates (the T-DNA insertion carried the NPT marker gene). Of the seedlings that germinated, 160 were kanamycin resistant and 107 were kanamycin sensitive. These data more closely fitted a 2:1 (chi-sq., 1df, = 5.5, $0.05 > P > 0.01$) than a 3:1 (chi-sq., 1df, = 32, $P < 0.001$) ratio. Such a segregation ratio suggested that a homozygous class of kanamycin resistant seedlings was absent from the progeny of KO.G374 plant. G374 can be a herbicide target.

Closely Related Genes from Other Species

Similar sequences to G374 are present in tomato and *Medicago truncatula*, and these sequences can be orthologs.

Example VIII: Identification of Homologous Sequences

Homologous sequences from *Arabidopsis* and plant species other than *Arabidopsis* were identified using database sequence search tools, such as the Basic Local Alignment Search Tool (BLAST) (Altschul et al. (1990) *J. Mol. Biol.* 215:403-410; and Altschul et al. (1997) *Nucl. Acid Res.* 25: 3389-3402). The tblastx sequence analysis programs were employed using the BLOSUM-62 scoring matrix (Henikoff, S. and Henikoff, J. G. (1992) *Proc. Natl. Acad. Sci. USA* 89: 10915-10919).

Identified non-*Arabidopsis* sequences homologous to the *Arabidopsis* sequences are provided in Table 5. The percent sequence identity among these sequences can be as low as 47%, or even lower sequence identity. The entire NCBI GenBank database was filtered for sequences from all plants except *Arabidopsis thaliana* by selecting all entries in the NCBI GenBank database associated with NCBI

taxonomic ID 33090 (Viridiplantae; all plants) and excluding entries associated with taxonomic ID 3701 (*Arabidopsis thaliana*). These sequences are compared to sequences representing genes of SEQ IDs NOs:2 - 2N, where $N = 2-561$, using the Washington University TBLASTX algorithm (version 2.0a19MP) at the default settings using gapped alignments with the filter "off". For each gene of SEQ IDs NOs:2 - 2N, where $N = 2-561$, individual comparisons were ordered by probability score (P-value), where the score reflects the probability that a particular alignment occurred by chance. For example, a score of $3.6e-40$ is 3.6×10^{-40} . In addition to P-values, comparisons were also scored by percentage identity. Percentage identity reflects the degree to which two segments of DNA or protein are identical over a particular length. Examples of sequences so identified are presented in Table 5. Homologous or orthologous sequences are readily identified and available in GenBank by Accession number (Table 5; Test sequence ID). The identified homologous polynucleotide and polypeptide sequences and homologues of the *Arabidopsis* polynucleotides and polypeptides may be orthologs of the *Arabidopsis* polynucleotides and polypeptides (TBD: to be determined).

Example IX Introduction of polynucleotides into dicotyledonous plants

SEQ ID NOs:1-(2N - 1), wherein $N = 2-561$, paralogous, orthologous, and homologous sequences recombined into pMEN20 or pMEN65 expression vectors are transformed into a plant for the purpose of modifying plant traits. The cloning vector may be introduced into a variety of cereal plants by means well-known in the art such as, for example, direct DNA transfer or *Agrobacterium tumefaciens*-mediated transformation. It is now routine to produce transgenic plants using most dicot plants (see Weissbach and Weissbach, (1989) *supra*; Gelvin et al., (1990) *supra*; Herrera-Estrella et al. (1983) *supra*; Bevan (1984) *supra*; and Klee (1985) *supra*). Methods for analysis of traits are routine in the art and examples are disclosed above.

Example X Transformation of Cereal Plants with an Expression Vector

Cereal plants such as corn, wheat, rice, sorghum or barley, may also be transformed with the present polynucleotide sequences in pMEN20 or pMEN65 expression vectors for the purpose of modifying plant traits. For example, pMEN020 may be modified to replace the NptII coding region with the BAR gene of *Streptomyces hygroscopicus* that confers resistance to phosphinothricin. The KpnI

and BglII sites of the Bar gene are removed by site-directed mutagenesis with silent codon changes.

The cloning vector may be introduced into a variety of cereal plants by means well-known in the art such as, for example, direct DNA transfer or *Agrobacterium tumefaciens*-mediated transformation. It is now routine to produce transgenic plants of most cereal crops (Vasil, I., Plant Molec. Biol. 25: 925-937 (1994)) such as corn, wheat, rice, sorghum (Cassas, A. et al., Proc. Natl. Acad Sci USA 90: 11212-11216 (1993) and barley (Wan, Y. and Lemeaux, P. Plant Physiol. 104:37-48 (1994). DNA transfer methods such as the microprojectile can be used for corn (Fromm, et al. Bio/Technology 8: 833-839 (1990); Gordon-Kamm et al. Plant Cell 2: 603-618 (1990); Ishida, Y., Nature Biotechnology 14:745-750 (1990)), wheat (Vasil, et al. Bio/Technology 10:667-674 (1992) ; Vasil et al., Bio/Technology 11:1553-1558 (1993); Weeks et al., Plant Physiol. 102:1077-1084 (1993)), rice (Christou Bio/Technology 9:957-962 (1991); Hiei et al. Plant J. 6:271-282 (1994); Aldemita and Hodges, Planta 199:612-617; Hiei et al., Plant Mol Biol. 35:205-18 (1997)). For most cereal plants, embryogenic cells derived from immature scutellum tissues are the preferred cellular targets for transformation (Hiei et al., Plant Mol Biol. 35:205-18 (1997); Vasil, Plant Molec. Biol. 25: 925-937 (1994)).

Vectors according to the present invention may be transformed into corn embryogenic cells derived from immature scutellar tissue by using microprojectile bombardment, with the A188XB73 genotype as the preferred genotype (Fromm, et al., Bio/Technology 8: 833-839 (1990); Gordon-Kamm et al., Plant Cell 2: 603-618 (1990)). After microprojectile bombardment the tissues are selected on phosphinothricin to identify the transgenic embryogenic cells (Gordon-Kamm et al., Plant Cell 2: 603-618 (1990)). Transgenic plants are regenerated by standard corn regeneration techniques (Fromm, et al., Bio/Technology 8: 833-839 (1990); Gordon-Kamm et al., Plant Cell 2: 603-618 (1990)).

The plasmids prepared as described above can also be used to produce transgenic wheat and rice plants (Christou, Bio/Technology 9:957-962 (1991); Hiei et al., Plant J. 6:271-282 (1994); Aldemita and Hodges, Planta 199:612-617 (1996); Hiei et al., Plant Mol Biol. 35:205-18 (1997)) that coordinately express genes of

interest by following standard transformation protocols known to those skilled in the art for rice and wheat Vasil, et al. Bio/Technology 10:667-674 (1992) ; Vasil et al., Bio/Technology 11:1553-1558 (1993); Weeks et al., Plant Physiol. 102:1077-1084 (1993)), where the bar gene is used as the selectable marker.

All references, publications, patent documents, web pages, and other documents cited or mentioned herein are hereby incorporated by reference in their entirety for all purposes. Although the invention has been described with reference to specific embodiments and examples, it should be understood that one of ordinary skill can make various modifications without departing from the spirit of the invention. The scope of the invention is not limited to the specific embodiments and examples provided.

We claim:

1. A transgenic plant comprising a recombinant polynucleotide having a nucleotide sequence selected from the group consisting of:
 - (a) a nucleotide sequence encoding a polypeptide comprising a sequence selected from those of SEQ ID NOs: 860, 802, 240, 274, 558, 24, 1120, 44, 460, 286, 120, 130, 134, 698, 832, 580, 612, and 48, or a complementary nucleotide sequence thereof;
 - (b) a nucleotide sequence of SEQ ID NOs: 859, 801, 239, 273, 557, 23, 1119, 43, 459, 285, 119, 129, 133, 697, 831, 579, 611, 47, or a complementary nucleotide sequence thereof; and
 - (c) a nucleotide sequence that hybridizes under stringent conditions to a nucleotide sequence of one or more polynucleotides of: (a) or (b).
2. The transgenic plant of claim 1 wherein the transgenic plant possesses an altered trait as compared to another plant, or the transgenic plant exhibits an altered phenotype as compared to another plant, or the transgenic plant expresses an altered level of one or more genes associated with a plant trait as compared to another plant, wherein the other plant does not comprise the recombinant polynucleotide.
3. The transgenic plant of claim 1 wherein the plant possesses an altered trait as compared to another plant wherein the trait is an alteration in one or more physical characteristics selected from the group consisting of: the number of trichomes, fruit and seed size and number, yield of stems, leaves, inflorescences, or roots, stability of seeds during storage, susceptibility of the seed to shattering, root hair length and quantity, internode distances, or the quality of seed coat.
4. The transgenic plant of claim 1 wherein the plant possesses an altered trait as compared to another plant wherein the trait is an alteration in a plant growth characteristic selected from the group consisting of: growth rate, germination rate of seeds, vigor of plants and seedlings, leaf and flower senescence, male sterility, apomixis, flowering time, flower abscission, rate of nitrogen uptake, osmotic sensitivity to soluble sugar concentrations, biomass or transpiration characteristics,

apical dominance, branching patterns, number of organs, organ identity, and organ shape or size.

5. The transgenic plant of claim 1 wherein the plant possesses an altered trait as compared to another plant wherein the trait is an alteration in one or more characteristics selected from the group consisting of protein or oil production, seed protein or oil production, insoluble sugar level, soluble sugar level, and starch composition.
6. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:860.
7. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:802.
8. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:240.
9. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:274.
10. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:558.
11. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:24.
12. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:1120.
13. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:44.

14. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:460.
15. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:286.
16. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:120.
17. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:130.
18. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:134.
19. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:698.
20. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:832.
21. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:580.
22. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:612.
23. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:48.
24. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:859.

25. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:801.
26. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:239.
27. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:273.
28. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:557.
29. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:23.
30. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:1119.
31. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:43.
32. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:459.
33. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:285.
34. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:119.
35. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:129.

36. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:133.
37. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:697.
38. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:831.
39. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:579.
40. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:611.
41. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:47.
42. The transgenic plant of claim 1, further comprising a constitutive, inducible, or tissue-specific promoter operably linked to said nucleotide sequence.
44. The transgenic plant of claim 1, wherein the plant is selected from the group consisting of: soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower, alfalfa, sugarcane, turf, banana, blackberry, blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, pumpkin, spinach, squash, sweet corn, tobacco, tomato, watermelon, mint and other labiates, rosaceous fruits, and vegetable brassicas.
44. The transgenic plant of claim 1 wherein the encoded polypeptide is expressed and regulates transcription of a gene.

45. A method of using the transgenic plant of claim 1 to grow a progeny plant from a parent plant, the method comprising crossing the transgenic plant with another plant, selecting seed, and growing the progeny plant from the seed.
46. An isolated or recombinant polynucleotide comprising a nucleotide sequence selected from the group consisting of:
- (a) a nucleotide sequence encoding a polypeptide comprising a sequence selected from SEQ ID NOs: 240, 274, 558, 286, 698, and 832, or a complementary nucleotide sequence thereof;
 - (b) a nucleotide sequence of SEQ ID NOs: 239, 273, 557, 285, 697, 831, or a complementary nucleotide sequence thereof; and
 - (c) a nucleotide sequence that hybridizes under stringent conditions to a nucleotide sequence of one or more of: (a) or (b).
47. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:240.
48. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:274.
49. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:558.
50. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:286.
51. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:698.

52. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:832.
53. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:239.
54. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:273.
55. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:557.
56. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:285.
57. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:697.
58. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:831.
59. The isolated or recombinant polynucleotide of claim 46, further comprising a constitutive, inducible, or tissue-specific promoter operably linked to the nucleotide sequence.
60. The isolated or recombinant polynucleotide of claim 46 wherein the encoded polypeptide is expressed and regulates transcription of a gene.
61. A vector comprising the isolated or recombinant polynucleotide of claim 46.
62. A host cell comprising the vector of claim 61.

63. A method of using the isolated or recombinant polynucleotide of claim 46 for producing a plant having a modified trait, the method comprising selecting a polynucleotide that encodes a polypeptide, inserting the polynucleotide into an expression vector, introducing the vector into a plant or a cell of a plant to overexpress the polypeptide, thereby producing a modified plant, and selecting a modified plant for a modified trait.
64. The method of claim 63 wherein the plant possesses a modified trait as compared to another plant wherein the trait is an alteration in one or more physical characteristics selected from the group consisting of: the number of trichomes, fruit and seed size and number, yield of stems, leaves, inflorescences, or roots, stability of seeds during storage, susceptibility of the seed to shattering, root hair length and quantity, internode distances, or the quality of seed coat.
65. The method of claim 63 wherein the plant possesses a modified as compared to another plant wherein the trait is an alteration in a plant growth characteristic selected from the group consisting of: growth rate, germination rate of seeds, vigor of plants and seedlings, leaf and flower senescence, male sterility, apomixis, flowering time, flower abscission, rate of nitrogen uptake, osmotic sensitivity to soluble sugar concentrations, biomass or transpiration characteristics, apical dominance, branching patterns, number of organs, organ identity, and organ shape or size.
66. The method of claim 63 wherein the plant possesses a modified trait as compared to another plant wherein the trait is an alteration in one or more characteristics selected from the group consisting of protein or oil production, seed protein or oil production, insoluble sugar level, soluble sugar level, and starch composition.
67. A modified plant produced by the method of claim 63.
68. A method of using the plant of claim 67 to grow a progeny plant from a parent plant, the method comprising crossing the transgenic plant with another plant, selecting seed, and growing the progeny plant from the seed.

69. The plant produced by the method of claim 68.

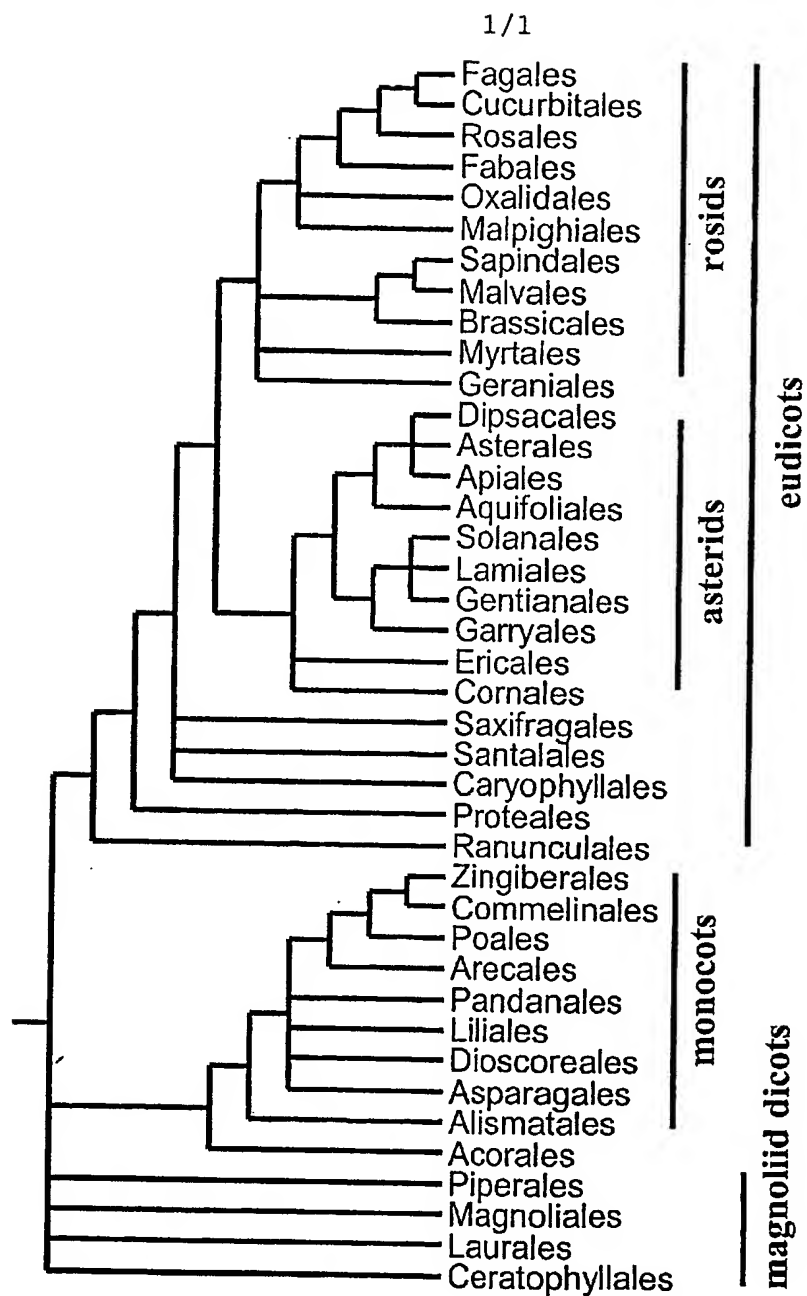


Figure 1

SEQUENCE LISTING

<110> Mendel Biotechnology, Inc.

Ratcliffe, Oliver

Riechmann, Jose Luis

Adam, Luc J.

Dubell, Arnold T.

Heard, Jacqueline E.

Pilgrim, Marsha L.

Jiang, Cai-Zhong

Reuber, T. Lynne

Creelman, Robert A.

Pineda, Omaira

Yu, Guo-Liang

Broun, Pierre E.

<120> YIELD-RELATED POLYNUCLEOTIDES AND
POLYPEPTIDES IN PLANTS

<130> 514442002041

<150> 60/310,847

<151> 2001-08-09

<150> 60/336,049

<151> 2001-11-19

<150> 60/338,692

<151> 2001-12-11

<150> 10/171,468

<151> 2002-06-14

>G1275 (58..579)

CCAAGAAAAGGGGAAGATCACGCATTCTTATAGGCGTAATTCGTAAATAGTGGTGAGTATG
AATGATGCAGACACAACTTGGGGAGTAGTTTCAGCGATGATACTCACTCTGTGTTCGAG
TTTCCGGAGCTAGACTTGTCTAGATGAATGGATGGATGATGATCTTGTGTCTGCGGTTTCC
GGGATGAATCAGTCTTATGGTTATCAGACTAGTGATGTTGCTGGTGCTTTATTCTCAGGT
TCTTCTAGCTGTTTCAGTCATCTGAATCTCCAAGTACCAAACTTATGTTGCTGCTACA
GCCACTGCTTCTGCGGACAACCAAAACAAGAAAGAAAAGAAAAAATTAAAGGGAGAGTT
GCGTTCAAGACACGGTCCGAGGTGGAAGTGCTTGACGACGGGTTCAAGTGGAGAAAGTAT
GGGAAGAAGATGGTGAAGAACAGCCACATCCAAGAACTACTACAAATGTTCAAGTTGAT
GGCTGTCCCGTGAAGAAAAGGGTTGAACGAGACAGAGATGATCCGAGCTTTGTGATAACA
ACTTACGAGGGTTCCCAATCACTCAAGCATGAACTAAGACTCGAACTAAGGCTCAAGG
CGACCATGCTATATTCAGCACATCTTATTTTCTATGGTTACGAACGATACTTAAACTGC
TTCTAGTTCTTTATATCCATTGTAAACTGGTTGCAGGTTCACAAATTTTGAGAGGTTTAT
GACATTCTAAATCTGTAGTACTTATATA

>G1275 Amino Acid Sequence (domain in AA coordinates: 113-169)
MNDADTNLGSSFDDTHSVFEFPELDLSDEWMDDLVSAVSGMNQSYGYQTS DVAGALFS
GSSSCFSPHESPSTKTYVAATATASADNQNKKEKKIKGRVAFKTRSEVEVLDDGFKWRK
YGKKMVKNSPHPRNYYKCSVDGCPVKRVERDRDDPSFVITTYEGSHNHSSMN*

>G1411 (110..856)

TAAAGAAAACCTGAACAACCTTAAAGTACTGTATAAATCCTATATCAAATTTTTTTTTTG
GAAGAAAAGGCTATATTTAAAGAAAATCAAGCAAAAGTAGATCCTCGGATGTATGGGAA
GAGGCCTTTTGGAGGTGATGAATCTGAAGAAAGGGAAGAAGATGAGAACTTGTTCCTGGT
CTTCTCGGCCCGATCTCAACACGACATGCGTGTTATGGTCTCGGCCTTGACTCAAGTAAT

CGGAAACCAACAAAGCAAATCTCATGATAACATCAGCTCTATTGATGATAACTATCCTTC
TGTGTATAATCCACAAGACCCTAATCAACAAGTTGCGCCTACTCATCAAGACCAAGGGGA
CTTGAGGAGGAGACATTATAGAGGTGTAAGGCAAAGGCCATGGGGAAAGTGGGCAGCTGA
AATCCGAGACCCAAAAAAGGCGGCACGTGTGTGGCTCGGGACATTTGAAACCGCTGAATC
TGCGGCCTTAGCTTATGATGAAGCAGCCCTAAAGTTCAAAGGAAGCAAAGCAAACCTCAA
TTTCCCGGAGAGGGTTTCACTTGGAAAGTAAGTCTACATATTACTCCTCCAACCAAATTCC
ACAAATGGAACCAAAAGTATACCGAACTATAATCAATACTATCATGATGCGAGTAGTGG
TGATATGCTAAGTTTTAATTTGGGCGGTGGGTATGGGAGTGGTACCGGATATTCAATGTC
TCATGATAATAGTACTACGACTGCTGCTACAACCTTCTCGTCTTCTGGTGGCTCTTCTAG
GCAACAAGAAGAGCAAGATTATGCCAGATTCTGGCGCTTTGGGGATTCTTCTTCTCTCC
TCATTCCGGGATATTAAATTAGGAGATTGATCAGTTACTTGTGATGAAGTAATGATACATT
TCCCGTCAAAATTGAGATGATCATATGCTTCTGAATGTTTTGAGTGTCAATTTTGTCT
TCCGCGTTAAGATTTATTGAACGTGTTTTCTTGTTTTTTGGTTAAAAA

>G1411 Amino Acid Sequence (domain in AA coordinates: 87-154)
MYGKRPFPGDESEEREEDENLFPVFSARSQHDMRVMVSALTQVIGNQSKSHDNISSIDD
NYPVYNPDQPNQVAPTHQDQGLRRRHRYGVRQRPWGKWAIEIRDPKKAARVWLGTPE
TAESAALAYDEAALKFKGSKAKLNFPERVQLGSNSTYYSSNQIPQMEPQSIIPNYNQYYHD
ASSGDMLSFNLGGYSGTGYSMSHDNSTTTAATSSSSGGSSRQQEEQDYARFWRFGDS
SSSPHSGY*

>G1488 (1..996)
ATGGAAGATGAAGCACATGAATTCTTCCACACATCTGATTTTGCCGTTGATGACCTTTTA
GTTGATTTCTCTAACGATGATGACGAAGAAAACGATGTTGTTGCTGATTCCACCACTACC
ACCACCATAACCGACAGCTCTAACTTCTCCGCTGCTGATCTTCCAGTTTCCACGGTGAT
GTTCAAGACGGCAGCTAGCTTCTCCGCTGACCTTTGTATACCTTCTGATGATTTGGCTGAT
GAGTTAGAGTGGCTTTCGAACATTGTGGATGAATCATTGTCGCTGAAGATGTACACAAG
CTCGAGCTAATATCCGGTTTTAAGAGTCCGACCGGACCCGAAATCCGATACCGGAAGCCCG
GAAACCCGAAATAGCAGCAGTCCGATTTTACTACCGAGCTTCTGTACCGGCCAAAGCT
AGAAGCAAACGCTCACGCGCCGTGCTAATTGGGCCCTCAGTGGGCTTCTCAAGGAA
ACGTTTTACGACAGTCTTTTACCAGGAGAAACATTCTCTAGCCAACAACACTTGTCT
CCGCCAACCTCGCCGCTTTGTTGATGGCTCCGCTAGGGAAAAAGCAAGCCGTTGATGGA
EGACACCGACGGAAGAAGGATGTTTCTTACCAGGAGTCTGGTGGCGCAGAGGAGAGACGG
TGTCTCCACTGCGCCACGGATAAGACTCCGCAATGGCGGACAGGCCCAATGGGCCCGAAG
ACGTTGTGCAACGCTTGCGGTGTTAGGTACAAATCGGGACGTTTAGTGCCGAGTATCGG
CCGCGCGGAGTCCGACGTTTGTGCTGGCGAAACACTCAAATTCTCATCGAAAGTTATG
GAGTCCGCGGACAGAAGGAGATGAGTAGGGCCCATCATGAGTTCATACATCACCATCAC
GGTACGGACACTGCCATGATTTTCGACGTTTCATCGGACGGTGATGATTACTTGATCCAC
CACAACGTTGGCCAGATTTTCCAGACGCTTATTGA

>G1488 Amino Acid Sequence (domain in AA coordinates: 221-246)
MEDEAHEFFHTSDFAVDDLVDVFNDDDEENDVADSTTTTTITDSSNFSAADLPSFHGD
VQDGTSPFSGDLICPSDDLDELEWLSNIVDESLSPELVHKLISGFKSRPDPKSDTGSP
ENPNSSSPIFTTDSVPAKARSKRSRAACNWAASRGLLKETFYDSPFTGETILSSQQHLS
PPTSPPLLMAPLGKKQAVDGGHRRKDVSSPESGGAEERRCLHCATDKTPQWRTGPMGPK
TLCNACGVRYKSGRLVPEYRPAASPTFVLAKHSNSHRKVMELRRQKEMSRAHHEFIHHHH
GTDAMIFDVSSDGDYLIHNVGPDFRQLI*

>G1499 (159..833)
TCGACTCCTTAATTGCATCACCAACCTAACCCCTCATCATATGATCACCTTTCTTTTTT
CCTTTAATATATATATATGCTCACACACACATATATATATACATATAAGCATCGCCTC
AAGCATTAATTTTTACGAACCAACAAACAAAAATTATGAATAATTATAATATGAACC
CATCTCTCTTCCAAAATTACACTTGGAAACAACATCATCAACAGCAGCAACAACAACA
AGAATGATGATCATCATCAACATAATAATGATCCAATCGGTATGGCCATGGACAGT
ACACACAGCTCCATATCTTCAATCCTTTCTCTTCTTCTCATTTCCCTCCTCTCTTCTT
CCCTCACAACCACTCTTCTCTCCGGAGATCAAGAAGACGACGAAGACGAAGAAGAAC
CTCTAGAGGAACTCGGTGCTATGAAGGAAATGATGTACAAGATCGCAGCCATGCAATCGG
TTGACATCGACCCAGCAACCGTCAAGAAACCCAAACGCCGTAAAGTGAAGGATCTCCGACG
ACCCTCAGAGTGTGGCGGCTAGACATCGCCGTGAGAGAATCAGTGAGAGGATCAGAATTC
TTCAGAGACTCGTGCCAGGTGGCACTAAATGGATACGGCTTCAATGCTCGATGAAGCTA

TACGCTATGTCAAGTTCTTGAAACGGCAGATCCGGCTACTCAATAATAATACCGGATATA
CTCCTCCGCCGCCGCAAGATCAAGCTTCTCAGGCGGTGACGACGTCATGGGTTTCACCGC
CACCACGCCCAAGTTTCGGCCGTGGGGGAAGAGGAGTAGGAGAATTAATCTAGACAAGAT
GACATTTCATTAGTAGTAATAAATTATGCTATAATGTGTGAGTAATGGTGCAATTATG
GA

>G1499 Amino Acid Sequence (domain in AA coordinates: 118-181)
MNNYNMNP SLFQNYTWNNIINSSNNNNKNDH HHQHNDPIGMAMDQYTLHIFNPFSSS
HFPPLSSSLT TTTLLSGDQEDDEDEEEPLEELGAMKEMMYKIAAMQSVDIDPATVKKPKR
RNVRISSDPQSVAARHRRERISERIRILQRLVPGGTKMDTASMLDEAIRYVKFLKQIRL
LNNNTGYTPPPPDQASQAVTTSWVSPPPPSFGRGGRGVGELI*

>G1543 (1..828)
ATGATAAACTACTATTACGTACATATGCACATACACATATAAACTATATGCTCTATAT
CATATGGATTACGCATGCGTGTGTATGTATAAAATAAAGGCATCGTCACGCTTCAAGTT
TGTCTCTTTTATATTAACTGAGAGTTTTCTCTCAAACCTTACCTTTCTTCTTCGATC
CTAGCTCTTAAGAACCCTAATAATTCATTGATCAAAATAATGGCGATTTTGCCGGAAGAAC
TCTTCAAACCTTGGATCTTACTATCTCCGTTCCAGGCTTCTCTTCATCCCCCTCTCTCCGAT
GAAGGAAGTGGCGGAGGAAGAGACCAGCTAAGGCTAGACATGAATCGGTTACCGTCGTCT
GAAGACGGAGACGATGAAGAATTCAAGTCACGATGATGGCTCTGCTCCTCCGCGAAAGAAA
CTCCGTCTAACCAGAGAACAGTCACGCTTCTTGAAGATAGTTTCAGACAGAATCATACC
CTTAATCCCAAACAAAAGGAAGTACTTGCCAAGCATTTGATGCTACGGCCAAGACAAATT
GAAGTTTGGTTTCAAACCGTAGAGCAAGGAGCAAATTGAAGCAAACCGAGATGGAATGC
GAGTATCTCAAAGGTGGTTTGGTTTCAATTAACGGAAGAAAACACAGGCTCCATAGAGAA
GTAGAAGAGCTTAGAGCCATAAAGGTTGGCCCAACAACGGTGAACCTCTGCCTCGAGCCTT
ACTATGTGTCTCGCTGCGAGCGAGTTACCCCTGCCGCGAGCCCTTCGAGGGCGGTGGTG
CCGCTTCCGGCTAAGAAAACGTTTCCGCCGCAAGAGCGTGATCGTTGA

>G1543 Amino Acid Sequence (domain in AA coordinates: 135-195)
MIKLLFTYICTYTYKLYALYHMDYACVMYKYKGI VTLQVCLFYIKLRVFLSNFTFSSSI
LALKNPNNSLIKIMAILPENSSNLDLTISVPGFSSSPLSDEGSGGGRDQLRLDMNRLPSS
EDGDDEEFSHDDGSAPPRKKRLRLTREQSRLLDSFRQNHTLNPQKEVLAKHMLRPRQI
EVWFQNRARRSKLKQTEMECEYLKRWFGSLTEENHRLHREVEELRAIKVGPTTVNSASSL
TMCPCRERVTPAASPSRAVVPVPAKKTFFPQERDR*

>G1635 (1..1164)
ATGGCGTCGTCTCCGTGACTGCAAATGTTCAAGGTACCAACGCTTCTTTGAGGAATAGA
GATGAAGAACTGCAGACAAGCAGATACAATTCAATGACCAAAGTTTGGGGGAAATGAC
TATGCACCCAAGGTACGGAAGCCATACCGATAACAAAAGAGAGAGAGATGGACAGAT
GAAGAGCACAGAAGTTTGTGTAAGCCTTGAAATTATACGGGCGAGCTTGAGAGACGAATA
GAAGAACATGTGGGCTCAAAGACCGCAGTTTCAAGATTGCAAGCCATGCTCAGAAGTTTTC
TCTAAGGTTGCTCGAGAAGCACTGGAGGTGATGGGAGCTCAGTAGAGCCGATTGTAATA
CCTCCTCCTCGTCCCAAGAGAAAGCCAGCATCCGTACCCCTCGTAAGTTTGGGAACGAG
GCAGATCAAACAAGTAGATCGGTTTCTCCCTCAGAACGTGATACTCAATCTCCAACCTCT
GTGTTGTCCACTGTGGATCAGAAGCATTGTGTTCCCTTGATTTCGAGTTTACCCAATCGA
AGCTTGTCCCCAGTTTCTTCTGCATCACCACCAGCTGCTCTTACAACCACTGCAAATGCA
CCTGAAGAGCTTGAGACTCTGAAGCTGGAGTTGTTTCTAGTGAGAGACTCTTAAACAGG
GAGAGCTCGATCAAGGAACCAACGAAGCAAAGTCTTAAACTCTTTGGGAAGACAGTTTG
GTATCTGATTCAGGCATGTCTCTTCTCTAACAACCTTCAACATATTGTAAATCCCCAATT
CAGCCATTACCACGGAACCTCTCATCATCAAGACACTACCCATAATAAGAACTCACAA
GAAGAACCTTTGAGCTGCTGGATACAAAGTCCCTCTTAAGCAAGAAGATGTGGAAAATAGA
TGTTTGGATTTCAGGAAAGGCTGTCCAAAACGAAGGATCATCGACTGGATCAAACACTGGT
TCGGTGGATGATACGGGACACACGGAAGACACAGAACCCGAAACAATGCTATGTCAA
TGGGAGTTTAAACCAAGTGAGAGGTCTGCATTTTCTGAGCTCAGAAGAACAACCTCCGAG
TCAAATTCAGAGGATTTGGTCCATACAAGAAGAGAAAGATGGTAACAGAAGAAGAAGAG
CATGAGATTCACTCTCACTTATAA

>G1635 Amino Acid Sequence (domain in AA coordinates: 44-104)
MASSPLTANVQGTNASLRNRDEETADKQIQFNDSFGGNDYAPKVRKPYTITKERERWTD
EEHKKFVEALKLYGRAWRRIEEHVSGKTAVQIRSHAQKFFSKVAREATGGDGSSVEPIVI
PPRPKRKPAHPYPRKFGNEADQTSRSVSPSERDTQSPTSVLSTVGSEALCSLDSSSPNR
SLSPVSSASPPAALTTTANAPEELET LKLELFP SERLLNRESSIKEPTKQSLKLFKTVL

VSDSGMSSSLTTSTYCKSPIQPLPRKLSSSKTLPIIRNSQEELLSCWIVPLKQEDVENR
CLDSGKAVQNEGSSTGSNTGSDVDDTGHTTEKTEPETMLCQWEFKPSERSAFSELRRTNSE
SNSRFGPGPYKKRKMVTEEBEHEIHLHL*

>G1794 (160..1335)

TCTTTCTTTCTTCTCTTTGCTCTGTTTCTGTTTCTCTCTCTCTCTCTACAGAGTT
TTCTTTCCCTCGAAGAAAAAGAATATTTTAAATTTAATTTTCTCTGCGTTTATAAGCTT
TAAGTTTCAGAGGAGGAGGATTTAGAAGGAGGGTTTGTATGTGTGCTTAAAAGTGGCA
AATCAGGAAGATAACGTTGGCAAAAAAGCCGAGTCTATTAGAGACGATGATCATCGGACG
TTATCTGAAATCGATCAATGGCTTTACTTATTTCGCAGCCGAGACGACCACCACGTCAT
AGCTTCCCCTACGCAGCAGCCGCCTCCATCGTCGTCGTCCTCATCTCTTATCTCAGGTTTC
AGTAGAGAGATGGAGATGTCTGCTATTGTCTCTGCTTTGACTCACGTTGTTGCTGGAAAT
GTTCTCAGCATCAACAAGGAGGCGGTGAAGGTAGCGGAGAAGGGACTTCGAATTCGTCCT
TCTTCTCGGGGCGAGAAAAGGAGGAGAGAGGTGGAGGAAGGTGGCGCCAAAGCGGTTAAG
GCAGCTAATACTTTGACGGTTGATCAATATTTCTCCGGTGGTAGCTCTACTTCTAAAGTG
AGAGAAGCTTCGAGTAACATGTCTAGGTCGGGGCCCAACATACGAGTATACAACTACGGCA
ACTGCTAGTAGCGAAACGTCGTCGTTTAGTGGGGACCAACCTCGGCGAAGATACAGAGGA
GTTAGACAAAGACCATGGGGAAAGTGGGCGGCTGAGATTCCGAGATCCATTTAAAGCAGCT
AGAGTTTGGCTCGGTACGTTTCGACAATGCTGAATCAGCAGCAAGAGCTTACGACGAAGCT
GCACTTCGGTTTAGAGGCAACAAAGCCAAACTCAACTTCCCTGAAAACGTCAAACTCGTT
AGACCTGCTTCAACCGAAGCACAACCTGTGCACCAACCGCTGCTCAAAGACCGACCCAG
TCAAGGAACCTCGGGTTCAACGACTACCTTTTGCCCATAGACCTGCTTCAATCAAAGC
GTTCAATTCGCAGCCGTTGATGCAATCATACAATCTTTGGATTATACGATCAAATGTCGTTT
CAACAACAGTTTCAGCAACATCATCAACAATCTTTGGATTATACGATCAAATGTCGTTT
CCGTTGCGTTTCGGTCACACTGGAGGTTCAATGATGCAATCTACGTCGTCATCATCATCT
CATTCTCGTCTCTGTTTTCCCGGCTGCTGTTTCAGCCGCCACCAGAATCAGCTAGCGAA
ACCGGTTATCTCCAGGATATACAATGGCCATCAGACAGACTAGTAATAACTACAATAAT
AGTCCATCCTCCTGATGACTTGCTTCATTTATTTGTTTCACTATAGAGTAATAGAAAAC
AGGAAAATGATTATATGTTATAGAGTTATTTTCCAAATATTATAGGGTTTAGGTTGTTT
GTATTGTTCTGCTTTTCATCCTCTCATGCTTTTTTCTTAATTTATTATTTTTGCATTA
TAATTTTCGTTTCATTGTAAACAACATTAAGAACCATGGAGAAAGGAAAAAAGAG
AG

>G1794 Amino Acid Sequence (domain in AA coordinates: TBD)

MCVLKVANQEDNVGKKAESIRDDHRTLSEIDQWLYLFAAEDDHHRHSFPTQQPPSSSS
SSLISGFSREMESAIVSALTHVAGNVPQHQQGGEGSGEGTSNSSSSSGQRRREVEE
GGAKAVKAANTLTVDQYFSGSSSTSKVREASSNMSGPGPTYEYTTTATASSETSSFSGDQ
PRRRYRGVRQRPWGKWAIEIRDPFKAARVWLGTFDNAESAARAYDEAALFRGNKAKLNF
PENVKLVRPASTEAPVHQTAARPTQSRNSGTTTTLLPIRPASNQSVHSQPLMQSYNLS
YSEMARQQQFQQHQQSLDLVDQMSFPLRFHTGGSMQSTSSSSSHSRPLFSPAAPQP
PPESASETGYLQDIQWPSDKTSNNYNNSSPS*

>G1839 (38..592)

ATCACAGTTATGTTTCCATTTCATGCGCTATAAAAACCATGCTCACTCCCTTTTGTTCTTC
ACACCATTTGCAGGAAAAAATGAATAGTTGTCAGTCTAATCCCACCAAAATGGATAATTC
AGAAAATGTTCTATTTAATGATCAAAACGAAAATTTACACTTGTGTCACCACACCCTTC
TTCTTCGTACTTGACAAGAGATCAAGAGCAGGATCATGGTCTCTGCTCTGCGACAAGT
GATATCTAACTCCGGAGCTGACGACGCGTCATCATCAAACTTGATCATCACAAGCGTTCC
GCCTCCAGACGCTGGCCCTTGCTCTCTGTTGGCGTCGCCGTTGCTACGGCTGCACATT
ACAACGGCCCGACCGAGAGGTAAAGAAGGAGAAGAAATACAAGGAGTAAGGAAAAAACC
ATCGGGTAAATGGGEGGCGGAGATATGGGATCCGAGATCGAAATCAAGGAGGTGGCTTGG
AACGTTTCTTACGGCGGAGATGGCGGCACAATCTTACAATGATGCGGCGGCTGAGTATCG
AGCAAGACGTGGTAAAACAAACGGAGAAGGAATTAAACGGCGGTGGAGATGACTGAGAAG
GACATGGTTCGGTGATCATACACGGCGAGGTGGAAATGTTATTTACTATTGAAAATAA
ATTATTTATTATAGAGGGAGATATTACTCTTTACGCTTTTCATTAGATTTATTTTTATAA
GTTTTAAAGTATTTTATTGTTATAAAAAA

>G1839 Amino Acid Sequence (domain in AA coordinates: TBD)

MLTPFCSSHLQEKMNCSQSNPTKMDNSENVLFDQENFTLVAPHPSSSYLTRDQEHEI
MVSALRQVISNSGADDASSNLIITSVPPPDAGPCPLCGVAGCYGCTLQRPHEVKKKK
YKGVRRKPSGKWAIEIWDPRSKSRRWLGTFLTAEMAAQSYNDAAEYRARRKTNKEGIK

RRWR*

>G2108 (35..694)

GAGAGAGAAACATTGATCTCTGAATATTGTGAACATGTTGAAATCAAGTAACAAGAGAAA
 AAGCAAAGAAGAGAAGAAGTTACAAGAAGGGAAGTACCTTGGAGTGAGGAGACGTCCATG
 GGAAGATATGCGAGCTGAAATCAGAAACCCCTTTTACTAAAGAAAGACATTGGCTTGAAC
 GTTTGATACAGCCGAAGAAGCTGCTTTTGCATATGACGTTGCTGCTCGATCCATCAGCGG
 CTCTCTAGCTACAACAACTTCTTCTACACTGAAAACACCTCTTTAGAAAGACATCCACA
 ACAGTCTTTGGAGCCTCATATGACTTGGGGATCTTCTAGTCTCTGTCTTCTTCAAGATCA
 GCCTTTTGAAAACAACCATTTTGTGCTGATCCTATCTCTTCTTCTTTTCTCAAAAACA
 AGAGTCTTCTACCAATCTCACTAACACTTTCTCACATTGTTATAATGATGGTGATCATGT
 TGGCCAAAGCAAAGAGATTTCTTTACCTAATGATATGTCAAACAGTTTATTCCGTCATCA
 GGACAAAGTCGGTGAACATGACAAATGCAGACCATATGAAGTTTGGCTCAGTTCTCAGCGA
 CGAACCTCTCTGCTTTGAGTATGACTACATTGGGAATTATCTTCAGAGTTTTCTCAAAGA
 TGTCAACGACGATGCTCCACAGTTTCTTATGTGAGCTTGATTACCGATCCTTCAATTTA
 TG

>G2108 Amino Acid Sequence (domain in AA coordinates: 18-85)

MLKSSNKRKSKKEKKLQEGKYLGVRRRPWGRYAAEIRNPFTKERHWLGTFTDTEEAFAFAY
 DVAARSISGLATTNFFYTENTSLEHRHPQOSLEPHMTWGSSSLCLLDQPFENNHFVADP
 ISSFSQKQESSNTLNTFHSYNDGDHVGQSKEISLPNDMSNSLFGHQDKVGEHDNADH
 MKFGSVLSDEPLCFEYDIIGNYLQSFLLKDVNDAPQFLM*

>G2291 (27..797)

GCTTTCTCACCTTTTATAAAATAGAAAATGGAAAACAGCTACACCGTTGATGGTCACCGTC
 TTCAATATTCGGTTCCGTTAAGCTCCATGCGATGAAACAGTCAAAACTCCGAAACTTACG
 GATTATCCAAAGAGTCGCCGTTGGTCTGCGATGCCCTTGTTCGAAACCAACACTACTTCAT
 TCGATATCTCTTCTCTTTTCTCGTTTAAACCCAAAACAGAACCCGAAAACACGCATCGTG
 TCATGGACGATTCCATCGCCGCCGTCGTGGGCGAAAACGTTCTTTTCCGTGATAAAAACA
 AAGTCTCTGATCACTTGACCAAAGAGGTGGTGTGAAGCGGGGGCGGAAGATGCCGAGA
 AGACCGGAGGATTCATGGGAGTGAGAAAACGGCCGTGGGGGAGATGGTCGGCGGAGATAA
 GAGACAGGATAGGGCGGTGCAGACACTGGTTAGGAACGTTTCGACACGGCGGAAGAGGCAG
 CGCGTGGTATGACGCGGCGGCGAGGAGGCTTAGAGGGACCAAAGCCAAGACCAATTTCTG
 TGATTCTCCGCTTTTCCCAAGGAAATAGCTCAGGCTCAGGAGGATAATAGGATGAGGC
 AGAAGCAGAAGAAGAAGAAGAAAAAGTGAGTGTGAGGAAGTGTGTTAAAGTCACAT
 CGGTTGCACAGTTGTTTCGATGATGCCAATTTTATAAATTCTTCTAGTATTAAAGGAAATG
 TGATTAGTCTATTGATAATCTTGAAAAATGGGTCTAGAGCTTGATTGAGTTAGGGT
 TGTTGTCTAGGAAGTGATAAAGCACTCGTAGTTAAGTAGTTGTAGTT

>G2291 Amino Acid Sequence (domain in AA coordinates: TBD)

MENSYTVDGHRQLQYSVPLSSMHETSQNSETYGLSKESPLVCMPLFETNTTSFDISSLSF
 NPKPEPENTHRVMDSDIAAVGVENVLFGDKNKVSDHLTKEGGVKRGKMPQKTGGFMGVR
 KRPWGRWSAEIRDRIGRHNLGTFTDTEEAARAYDAAARRLRGTAKTNFVIPPLFPKE
 IAQAQEDNRMROKQKQKKKKKVSVRKCVKVTVAQLFDDANFINSSSIKGNVISSIDNLE
 KMGLELDLSLGLLSRK*

>G2452 (1..804)

ATGTCATCGTCGACGATGTACAGAGGAGTTAATATGTTTTTACCGGCAAACACAAACTGG
 ATTTTTCAAGAAGTCAGAGAAGCCACGTGGACGGCGGAGGAGAACAACGGTTCGAGAAA
 GCTCTCGCTTATCTGGACGACAAAGACAATCTTGAGAGCTGGTCCAAGATCGCAGATTTG
 ATTCCCGGCAAAACAGTAGCTGACGTCATTAAACGATACAAGGAGCTAGAGGATGATGTC
 AGCGACATCGAAGCCGGACTTATCCCCATTCCGGGATACGGCGGCGACGCTCCTCCGCT
 GCAAAACAGTGACTATTCTTTGGTCTAGAAAATCCAGCTACGGTTATGATTACGTCGTT
 GGAGGAAAGAGGAGTTTCGCCGGCGATGACTGATTGTTTAGGTCTCCGATGCCGGAAG
 GAGAGGAAGAAAGGAGTTCCGTGGACCGAGGACGAACACCTACGATTTCTGATGGGTTTG
 AAGAAATATGGAAGAGGAGATTGGAGAAACATAGCAAAAAGCTTTGTGACGACTCGAACG
 CCGACGCAAGTCGCTTACACGCTCAGAAATATTTCTTCGACAACCTCACAGATGGTAAA
 GACAAAAGACGATCAAGTATTCAGATATCACCCTGTAAACATCCCTGACGCAGACGCA
 TCCGCAACCGCCACGACCGCTGACGTAGCACTCTCTCTACTCCAGCCAATCTTTTGAC
 GTTTTCTTTCAGCCAAATCCTCATTACAGTTTCGCGTCTGCGTCTGAGTATTAT
 AATGCGTTTCCGCAGTGGAGTTAA

>G2452 Amino Acid Sequence (conserved domain in AA coordinates: 27-213)

>G2509 (143..934)

>G2509 Amino Acid Sequence (domain in aa coordinates: 89-156)

>G390 (1..2526)

6

GGAGGTCATGCTTACTCACCTGAAGACATGGGCTTATCCCGGGATATGTATTTACTGCAG
CTTTGTAGCGGCGTTGATGAAAATGTGGTTGGAGGTTGTGCTCAGCTTGTCTTTGCCCCA
ATCGATGAATCATTGCTGATGATGCACCTTTGCTTCCTTCTGGTTTCCGTGTCATACCA
CTCGACCAAAAAACAAATCCGAATGATCATCAATCTGCAAGTCGAACACGGGATCTAGCA
TCGTCCCTAGATGGTTCCACCAAAACCGATTCCGGAACAACTCTAGATTGGTCTTAACA
ATAGCCTTCCAGTTACGTTTGTGATAACCATTCCAGAGACAATGTTGCTACAATGGCGAGA
CAGTATGTGAGGAACGTTGTTGGTTTCGATTGAGAGAGTGGCTCTAGCCATTACGCCCTCGT
CCTGGCTCAATGCAACTTCCCCTTCCCCTGAAGCTCTCACTCTTGTCCGTTGGATCACC
CGTAGTTACAGTATTCATACAGGTGCAGATCTGTTTGGAGCTGATTCTCAGTCCTGTGGA
GGAGACACATTGCTTAAGCAACTCTGGGACCATAGTGATGCCATATTGTGCTGCTCCCTG
AAAATAATGCCTCACCGGTATTACATTTGCAAACCAAGCTGGTTTAGACATGCTTGAA
ACTACACTTGTGGCACTTCAGGATATAATGCTCGACAAAACACTTGATGACTCTGGTCTGT
AGAGCTCTTTGCTCCGAGTTCGCCAAGATCATGCAGCAGGGATATGCGAATCTTCCGGCA
GGAATATGTGTGTCGAGCATGGGCAGACCGGTTTCGTATGAGCAAGCGACGGTGTGGAAA
GTTGTTGATGACAACGAATCAAACCACTGCTTGGCTTTTACCCTCGTTAGTTGGTCTGTTT
GTTTGA

>G390 Amino Acid Sequence (domain in AA coordinates: 18-81)
MMAHHSMDRRDSPDKGFDGSGKYVRYTPEQVEALERVYAECPKPSSLRRQLIRECPILCN
IEPRQIKVWFQNRRCREKQKESARLQTVNRKLSAMNKLMEENDRLQKQVSNLVYENG
MKHRIHTASGTTTNDSCSESVVVGQQRQQNPTHQHPQRDVNNPANLLSIAEETLAEFLC
KATGTAVDWVQMIGMKPGPDSIGIVAVSRNCSGIAARACGLVSLPEPMKVAEILKDRPSWF
RDCRCVETLNVIPITGNGGTIELVNTQIYAPTTLAAARDFWTLRYSTSLDGSYVVCERSL
TSATGGPNGPLSSSFVRKMLSSGFLIRPCDGGGSIHIVDHDLDVSSVPEVLRPLYES
SKILAQKMTVAALHRVQIAQETSSEVQYSGGRQPAVLRTFSQRLCRGFNDVNGFVDDG
WSPMSSDGGEDITIMINSSAKFAGSQYGSFLPSFGSGVLCASMLLQNVPLVLIRF
LREHRAEWADYGVDAASAALRATPYAVPCVRTGGFPPSNQVILPLAQTLHEEFLEVRL
GGHAYSPEDMGLSRDMYLLQLCSGVDENVVGGCAQLVFAPIDESFADDAPLLPSGFRVIP
LDQKTNPNHDQSASRTRDLASSLDGSTKTDSETNSRLVLTIAFQFTFDNHSRDNVATMAR
QYVRNVVGSIQRVALAITPRPGSMQLPTSPEALTLVRWITRSYSIHTGADLFGADSQSCG
GDTLLKQLWDHSDAILCCSLKTNASPVFTFANQAGLDMLETTLVALQDIMLDKTLDDSGR
RALCSEFAKIMQQGYANLPAGICVSSMGRPVSYEQATVWKVVDNENHCLAFTLVSWSF
V*

>G391 (1..2559)

ATGATGATGGTCCATTTCGATGAGCAGAGATATGATGAACAGAGAGTCGCCGATAAAGGG
TTAGATTCCGGCAAGTATGTGAGGTACACGCCGGAGCAAGTGGAAGCTCTCGAGAGAGTT
TACACTGAGTGTCTTAAGCCAAGTTCTCTAAGAAGACAACAACCTCATACGTGAATGTCCG
ATTCTCTCTAACATCGAGCCTAAGCAGATCAAAGTTTGGTTTCAGAACCGCAGATGTCTGT
GAGAAGCAGAGGAAAGAGCTGCTCGTCTTCAAACAGTGAACAGAAAACCTCAATGCCATG
AACAACTCTTGATGGAAGAGAATGATCGTTTGCAGAAGCAAGTTTCTAACTTGGTCTAT
GAGAATGGCCACATGAAACATCAACTTCACACTGCTTCTGGGACGACCACAGACAACAGC
TGTGAGTCTGTGCTGCTGAGTGGTTCAGCAACATCAACAGCAAAACCCAAATCCTCAGCAT
CAGCAACGTGATGCTAACAAACCCAGCAGGACTCCTTTCTATAGCAGAGGAGGCCCTAGCA
GAGTTCTTTTCCAAGGCTACAGGAAGTCTGTTGACTGGGTTTCAGATGATTGGGATGAAG
CCTGGTCCGGATTCTATTGGCATAGTCGCTATTTTCGCGCAACTGCAGCGGAATTGCAGCA
CGTGCTGCGGCCCTCGTGAGTTTAGAACCCATGAAGGTTGCTGAAATTCTCAAAGATCGT
CCATCTTGGCTCCGAGATTGTGCAAGTGTGGATACTCTGAGTGTGATACCTGCTGGAAAC
GGTGGGACGATCGAGCTTATTTACACGCAGATGTATGCTCCTACGACTTTAGCAGCAGCT
CGTGACTTTTGGACGCTGAGATATAGCACATGTTTGAAGATGGAAGCTATGTGGTTTGT
GAAAGGTCGCTTACTTCTGCAACTGGTGGCCCCACTGGGCCACCTTCTTCAAACCTTTGTG
AGAGCTGAAATGAAACCAAGCGGTTTCTCATCCGTCCTTGGCATGGTGGTGGTTCCATT
CTCCACATTGTTGATCATGTTGATCTGGATGCTTGGAGTGTCCCTGAAGTCATGAGGCCT
CTCTATGAATCATCGAAGATTCTTGCTCAGAAAATGACTGTTGCTGCTTTGAGACATGTA
AGACAAATTGCACAAGAAACAAGTGGAGAAGTTCAGTATGGTGGAGGGCGCCAACCTGCCG
GTTTAAAGAACCTTCAGTCAAAGACTCTGTGCGGGTTTCAATGATGCTGTTAATGGTTTT
GTGGATGATGGATGTTACCAATGGGTAGCGATGGTGCAGAGGATGTTACTGTAATGATA
AACTTGTCCCTTGGGAAGTTTGGTGGGTCTCAGTACGGTAATTCTTCCCTTCAAGCTTT
GGTAGTGGCGTGCTTTGTGCCAAGGCATCTATGTTGCTTCAGAACGTTCCACCCGCTGTG

CTGGTTCGATTTCCTTAGAGAACACCGATCTGAATGGGCTGATTATGGCGTGGATGCTTAT
GCTGCTGCATCGCTCAGAGCAAGTCCTTTTGTCTGTTCTTGTGCTAGAGCTGGGGGGTTC
CCAAGTAACCAAGTCATTCTTCTCTTGCAGACAGTTGAACATGAAGAGTCACTTGAG
GTGGTTAGACTTGAAGGTACGCTTACTACCCGAAGACATGGGTTTAGCTCGGGATATG
TATTTGCTACAGCTTTGTAGCGGTGTTGATGAAAATGTGGTTGGAGGTTGTGCACAGCTT
GTATTTGCCCTATCGATGAATCATTTGCTGATGATGCACCTTTGCTTCCTTCCGGTTTC
CGCATCATACCTCTTGAACAGAAATCTACTCCGAACGGTGCATCTGCAAACCGTACCCTG
GATTTAGCCTCAGCTTTAGAAGGATCCACAGTCAAGCTGGTGAAGCCGACCCAAATGGC
TGTAACTTTAGGTCGGTACTAACCATAGCATTCCAGTTCACATTTGATAACCATTCAAGA
GACAGTGTGCTTCAATGGCACGTACGTACGTGCGAAGCATAGTAGGATCGATTAGAGG
GTTGCTCTAGCCATTGCTCCTCGTCCGCTGCTCCAATATCAGTCCAATATCTGTTCCCACT
TCCCCTGAAGCTCTCACTCTGGTCCGTTGGATCTCCCGGAGTTACAGCCTTCACACTGGT
GCAGATCTCTTTGGATCTGATTCTCAAACAGTGGTGACACGTTGCTGCATCAACTCTGG
AATCACTCTGATGCAATCTTGTGCTGCTCCCTCAAACAAACGCTTCACCGGTTTTCACA
TTCGCAAACCAAACCGGTTTAGACATGCTGGAACGACTCTTGTAGCCCTTCAAGACATA
ATGCTAGACAAGACCCCTTGACGAACCTGGTCGTAAGCTCTTGTCTGAGTTCCCAAG
ATCATGCAACAGGGCTATGCTCATCTGCCGGCAGGAGTATGTGCGTCAAGCATGGGAAGG
ATGGTATCTTACGAGCAGGCAACGGTGTGGAAGTTCTTGAAGACGATGAATCAAACAC
TGCTTAGCTTTCATGTTCTGTAATTGGTCTGTTCTGTTGA

>G391 Amino Acid Sequence (domain in AA coordinates: 25-85)
MMVHSMRDMNRESPPDKGLDSGKYVRYTPEQVEALERVYTECPKPSSLRRQQLIRECP
ILSNIEPKQIKVWFQNRRCREKQKQKEAARLQTVNRKLNAMNKLMEENDRLQKQVSNLVY
ENGHMKHQLHTASGTTTNDSCSVVVSQGHQHQNNPNPQHQQORDANNPAGLLSIAEEALA
EFLSKATGTAVDWVQMIGMKPGPDSIGIVAISRNCSGIAARACGLVSLPEMKVAEILKDR
PSWLRCRSDVTLSPVAPAGNGGTIELIYTQMYAPTTLAAARDFWTLRYSTCLEGDSYVVC
ERSLTSATGGPTGPPSSNFVRAEMKPSGFLIRPCDGGGSILHIVDHVDLDAWSVPEVMRP
LYESSKILAQKMTVAALRHVRQIAQETSSEVQYGGGRQPAVLRTFSQRLCRGFNDVNGF
VDDGWSPPMGSDGAEDVTVMINLSPGKFGGSQYGNFSLPSFGSGVLCAKASMLLQNVPPAV
LVRFLREHRSEWADYGVDAYAAASLRASPFAPPCARAGGFPSNQVILPLAQTVHEESLE
VVRLEGHAYSPEDMGLARDMYLLQLCSGVDENVVGGCAQLVFAPIDESFADDAPLLPSGF
RIIPLEQKSTPNGASANRTLDLASALEGSTRQAGEADPNCGNFRSVLTIAFQFTFDNHSR
DSVASMARQYVRSIVGSIQRVALAIAPRPGSNISPISVPTSPEALTLRVWISRSYSLHTG
ADLFGSDSQTSGDTLLHQLWNHSDAILCCSLKTNASPVFTFANQTGLDMLETTLVALQDI
MLDKTLDEPGRKALCSEFPKIMQQGYAHLPAVCASSMGRMVSYEQATVWKVLEDDSNH
CLAFMFVNWSFV*

>G438 (188..2716)
CGGGGTACCCAAGCCACGACCGTAGAATCTTCTTTTGTCTGAAAAGAATTACAATTTACG
TTTCTCTTACGATACGACGGACTTTCCGAAGAAATTAATTTAAAGAGAAAAGAAGAA
GCCAAAGAAGAAGAAGCTAGAAGAAACAGTAAAGTTTGAAGCTTTTGTAGGGTTCG
AGCTAAAATGGAGATGGCGGTGGCTAACCACCGTGAGAGAAGCAGTGACAGTATGAATAG
ACATTAGATAGTAGCGGTAAGTACGTTAGGTACACAGCTGAGCAAGTCGAGGCTCTTGA
GCGTGTCTACGCTGAGTGTCTTAAGCCTAGCTCTCTCCGTGACAAACAATTGATCCGTGA
ATGTTCCATTTTGGCCAATATTGAGCCTAAGCAGATCAAAGTCTGGTTTCAGAACCAGCAG
GTGTCGAGATAAGCAGAGGAAAGAGGCGTCGAGGCTCCAGAGCGTAAACCGGAAGCTCTC
TGCGATGAATAAACTGTTGATGGAGGAGAATGATAGGTTGCAGAAGCAGGTTTCTCAGCT
TGTCTGCGAAAATGGATATATGAAACAGCAGCTAACTACTGTTGTTAACGATCCAAGCTG
TGAATCTGTGGTCACAACCTCTCAGCATTCTGCTTAGAGATGCGAATAGTCTGCTGGATT
GCTCTCAATCGCAGAGGAGACTTTGGCAGAGTTCTATCCAAGGCTACAGGAAGTCTGT
TGATTGGGTTTCAATGCTGGGATGAAGCCTGGTCCGGATTCCGTTGGCATCTTTGCCAT
TTCGCAAAGATGCAATGGAGTGGCAGCTCGAGCCTGTGGTCTTGTAGCTTAGAACCTAT
GAAGATTGCAGAGATCCTCAAAGATCGGCCATCTTGGTTCCGTGACTGTAGGAGCCTTGA
AGTTTTCATATGTTCCCGGCTGGTAATGGTGGCACAATCGAGCTTGTATTATATGAGAC
GTATGCACCAACGACTCTGGCTCCTGCCCGGATTTCTGGACCCTGAGATACACACGAG
CCTCGACAATGGGAGTTTGTGGTTTGTGAGAGGTCGCTATCTGGCTCTGGAGCTGGGCC
TAATGCTGCTTCAGCTTCTCAGTTTGTGAGAGCAGAAATGCTTTCTAGTGGGTATTAAAT
AAGGCCTTGTGATGGTGGTGGTTCTATTATTCACATTGTGATCACCTTAATCTTGAGGC
TTGGAGTGTTCGGATGTGCTTCGACCCCTTTATGAGTCATCCAAGTCGTTGCACAAAA

AATGACCATTTCGCGTTGCGGTATATCAGGCAATTAGCCCAAGAGTCTAATGGTGAAGT
 AGTGATGGATTAGGAAGGCAGCCTGCTGTTCTTAGAACCTTTAGCCAAAGATTAAAGCAG
 GGGCTTCAATGATGCGGTTAATGGGTTTGGTGACGACGGGTGGTCTACGATGCATTGTGA
 TGGAGCGGAAGATATTATCGTTGCTATTAACTCTACAAAGCATTTGAATAATATTCTAA
 TTCTCTTTTCGTTCTTGGAGGCGTGCTCTGTGCCAAGGCTTCAATGCTTCTCCAAAATGT
 TCCTCTGCGGTTTTGATCCGGTTCCTTAGAGAGCATCGATCTGAGTGGGCTGATTTCAA
 TGTGATGCATATTCGCGTGTACACTTAAAGCTGGTAGCTTTGCTTATCCGGGAATGAG
 ACCAACAAGATTCAGTGGGAGTCAGATCATAATGCCACTAGGACATACAATTGAACACGA
 AGAAATGCTAGAAGTTGTTAGACTGGAAGGTCATTCTCTTGTCTCAAGAAGATGCATTAT
 GTCACGGGATGTCCATCTCCTTCAGATTTGTACCGGGATTGACGAGAATGCCGTTGGAGC
 TTGTTCTGAAGTGATATTTGCTCCGATTAATGAGATGTTCCCGGATGATGCTCCACTGT
 TCCCTCTGGATTCCGAGTCATACCCGTTGATGCTAAAACGGGAGATGTACAAGATCTGTT
 AACCGCTAATCACCGTACACTAGACTTAACTTCTAGCCTTGAAGTCGGTCCATCACCTGA
 GAATGCTTCTGGAACTCTTTTTCTAGCTCAAGCTCGAGATGTATTCTCACTATCGCGTT
 TCAATTCCTTTTGAAAACAACCTTGCAAGAAAATGTTGCTGGTATGGCTTGTCAGTATGT
 GAGGAGCGTGATCTCATAGTTCAACGTGTTGCAATGGCGATCTCACCGTCTGGGATAAG
 CCCGAGTCTGGGCTCCAAATTGTCCTCCAGGATCTCCTGAAGCTGTTACTCTTGTCTAGTG
 GATCTCTCAAAGTTACAGTCATCACTTAGGCTCGGAGTTGCTGACGATTGATTCACTTGG
 AAGCGACGACTCGGTACTAAAACCTTCTATGGGATCACCAAGATGCCATCCTGTGTTGCTC
 ATTAAAGCCACAGCCAGTGTTTATGTTTGCAGAACCAAGCTGGTCTAGACATGCTAGAGAC
 AACACTTGTAGCCTTACAAGATATAACACTCGAAAAGATATTCGATGAATCGGGTCGTAA
 GGCTATCTGTTCCGACTTCGCCAAGCTAATGCAACAGGGATTGCTTGCTTGCCCTCAGG
 AATCTGTGTGTCAACGATGGGAAGACATGTGAGTTATGAACAAGCTGTTGCTTGGAAAGT
 GTTGTGCTGCATCTGAAGAAAACAACAATCTGCATTGTCTGCCTTCTCCTTTGTAAA
 CTGGTCTTTTGTGTGATTGATTGACAGAAAAGACTAATTTAAATTTACGTTAGAGAAC
 TCAAAATTTTGGTTGTTTGGTTAGGTGTCTCTGTTTTGTTTTTAAAATTATTTTGATCAA
 A

>G438 Amino Acid Sequence (domain in AA coordinates: 22-85)
 MEMAVANHRERSSDSMNRHLDSGKYVRYTAEQVEALERVYAECPKPSLLRRQQLIRECS
 ILANIEPKQIKVWFQNRRCRDKQRKEASRLQSVNRKLSAMNKLMEENDRLQKQVSQILVC
 ENGYMKQQLTTVVNDPSCSVVTTPOHSLRDANSPAGLLSIAETLAEFLSKATGTAVDW
 VQMPGMPGPDSSVGIFAISQRCNGVAARACGLVSLPEPMKIAEILKDRPSWFRDCRSLEVF
 TMFPAGNGGTIELVYMQTYAPTTLAPARDFWTLRYTTSLDNCSFVVCERSLSGSGAGPNA
 ASASQFVRAEMLSSGYLIRPCDGGGSI IHIVDHLNLEAWSVPDVLRLPLYESSKVVAQKMT
 ISALRYIRQLAQESNGEVVYGLGRQPAVLRTFSQRLSRGFNDVNGFGDDGWSTMHCDGA
 EDIIIVAINSTKHLNINISNLSFLGGVLCASMLLQNVPPAVLIRFLREHRSEWADFNV
 AYSATLKGSAFAYPGMRPTRFTGSQIIMPLGHTIEHEEMLEVVRLEGHSLAQEDAFMSR
 DVHLLQICTGIDENAVGACSELI FAPINEMFPDDAPLVPSGFRVIPVDAKTGDVQDLLTA
 NHRTLDLTSSLEVGPSPENASGNSFSSSSSRCILTI AFQFPFENNLQENVAGMACQVRS
 VISSVQRVAMAI SPGISPSLGLSKLSPGSPEAVTLAQWISQSYSHHLGSELLTIDSLGSD
 DSVLKLWDHQDAILCCSLKPQPVFMFANQAGLDMLETTLVALQDITLEKIFDESGRKAI
 CSDFAKLMQQGFACLP SGICVSTMGRHVS YEQA VAWKVFAASEENNNNLHCLAFS FVNWS
 FV*

>G47 (38..472)
 CTTCTTCTTCACATCGATCATCATACAACAACAAAAATGGATTACAGAGAATCCACCGG
 TGAAAGTCAGTCAAAGTACAAAGGAATCCGTCGTCGGAAATGGGGCAAATGGGTATCAGA
 GATTAGAGTTCCGGGAACCTCGTGACCGTCTCTGGTTAGGTTCACTCTCAACAGCAGAAGG
 TGCCGCCGTAGCACACGACGTTGCTTTCTTCTGTTTACACCAACCTGATTCTTTAGAATC
 TCTCAATTTCCCTCATTTGCTTAATCCTTCACTCGTTTCCAGAACTTCTCCGAGATCTAT
 CCAGCAAGCTGCTTCTAAGCCCGGCATGGCCATTGACGCCGAATCGTCCACAGTACCAG
 CGTGAACCTGGATCGGAGATACGACGACGCTATTACGAGAATGGAGCTGATCAAGTGGA
 GCCGTTGAATATTTCACTGTATGATTATCTGGGCGGCCACGATCACGTTTGATTATCTC
 GACGGTCATGATCACGTTTGATCTTCTTTTGTAGTAAGATTTTGTACCATAATCAAAACAG
 GTGTGGTGCTAAAACTTACTCAAAACAAGATTAGGTACCACAGAGAAACAATCAAATGG
 TTGTGAATATACATTATAAGGTTTTGATTAAATGTTTGTCTGATTTAGTGAAGTTTG
 GTCCATTGTATACAAATCTATTCAAGAAACCTAGCGCGAGATCATGTTTCGTGATTGAAG
 ATTGAGATTTTTAAGTATTTCGTAATATTTTTGTAAAATACAAATAAAAAAAAAAAAAA

AAAAA

>G47 Amino Acid Sequence (domain in AA coordinates: 11-80)

MDYRESTGESQSKYKGIRRRKWKWVSEIRVPGTRDLWLGSFSTAEGAAVAHDVAFFCL
HQPDSLESLSNFPHLLNPSLVSRTPRSIQQAASNAGMAIDAGIVHSTSVNSGCGDITTTY
ENGADQVEPLNISVYDYLGGHDHV*

>G559 (89..1285)

aaagttgctagctttaatttgcgaacttactattcttatgtgtaataatcgtttgcaagg
tcgttgatttggtgataagtcagtagaaATGgataaggagaaatctccagcacctccttg
tggaggtcttctcctccatctccatcaggtcgatgctctgcattctcagaagctggtcc
cattggtcatggttcagatgctaatacgaatgagtcgatattagccgatgcttgataa
cccacctaagaagattggacatcggcgagctcattctgaaatacttactctccctgatga
tttgagctttgatagtgatcttgggtgtggttggaatgctgctgatggagcttcttctc
tgatgagactgaagaagatttgctctctatgtatcttgatggataagtttaattcttc
tgctacatcttctgcgaagttggtagccatcaggaactgcttggaataatgagacaat
gatgcagacaggcacaggtcaacttccaatcctcagaatacgggttaatagctctggcga
aaggccaagaatcaggcatcaacatagccaatctatggatggttcaatgaatatcaatga
gatgcttatgtcgggaatgaagatgattctgctattgatgctaagaagtctatgtctgc
tactaaacttgctgagcttgctctcattgatcctaaacgtgctaagaggatatgggcaaa
caggcagtcgcgagcagcatcaaaagaaaggaagacgagatacatatttgagcttgagag
aaaagtacagactttgcaaacagaggctacaactctctcagcccagttgacctcttaca
gagagacacaaatggcttgactgttgaaaacaatgagctgaagctgcggttacaacaat
ggagcagcaggttcacttgaggatgaactaaacgaagcactaaaggaggaaatccagca
tctgaaggtgttgactggccaagttgctccatcagcgttgaaactatgggtcggttgatc
aaaccagcagcaattctattccaacaatcagtcgaatgcaacaatcttagctgcaaaaca
gttccagcaacttcagattcattcacagaagcagcaacaacaacaacaacaacaaca
gcaacaccaacagcagcagcagcaacagcaacagtcagtttcaacagcaacagatgca
acagcttatgcagcagcggcttcaacagcaagaacaacaaatggagtaagactcaagcc
ttcacaagcccagaaagagaactGAggaatatgaatatgtcccacgtaagtgaagaggttc
tccttctgaacaattccttctcattcataaattgttggttcacatcacttgagctctc
ttggatttttaggggttttagctaacaca

>G559 Amino Acid Sequence (domain in AA coordinates: 203-264)

MDKEKSPAPPCGGLPPSPSGRCSAFSEAGPIGHGSDANRMSHDISRMLDNPPKKIGHRR
AHSEILTLDDLSFSDSLGVVGNADGASFSDTEEDLLSMYLDMDKFNSSATSSAQVGE
PSGTAWKNETMMQTGTGSTSNPQNTVNSLGERPRIHQHSQSMGSMNINEMLSGNEDD
SAIDAKKSMSATKLALALIDPKRAKRIWANRQSAARSKERKTRYIFELERKVQLQTEA
TTLSAQLTLQDRTNGLTVENNELKRLRLQTMEEQVHLQDELNEALKEEIQHLKVLTGQVA
PSALNYGSFSGSNQQQFYNNQSMQITLAAKQFQQLQIHSQKQKQQQQQQQQQHQHQHQHQ
QQYQFQQQQMQQLMQRLQQEQQNGVRLKPSQAQKEN*

>G568 (141..995)

GACCGGCTAAAGTCAAGAACCTCTCTCTGAGCTCTCACCACCTTCTCTCTACTCCCTC
TCTGCGTGTAGGATACTACTAGACAATTGACAACCAAGACTAAAGCTGTGTGTTGGTT
CACTTCTGTTCTCTTTTCCAATGTGTGTCATCAGCTAAGCATCAGAGAAACCATAGACTCT
CTGCTACAAACAAGAACCAGACTCTCACCAGTTTCTTCCATTTTCCTCATCACCAT
CGTCTTCTTCTTCATCATCATCAACCTCATCATCTCTCTTTTACCTTCTCAAGACTCTC
AAGCCCAGAAGAGATCTCTTGTCAACATGGAAGAAGTTTGGAAATGACATCAACCTTGCTT
CCATCCACCACCTAAACCCGACACAGCCCTCATCCACAACACAACCACGAGCCAAGTTCA
GGGGCCAAAACCACCACAACCAAAACCCTAACTCAATCTTCCAAGATTTTCTCAAAGGAT
CTTTGAACCAGGAAGCAGCACCACCAAGCCAGACCACGGGTTCTGCGCCTAATGGCGATT
CCACCACGGTCACTGTTCTTTACAGCTCTCTTTTCCACCTCCTGCAACTGTTCTGAGCT
TGAATTCGGGCGTGGCTTCGAGTTTCTCGATAACCAAGATCCTCTGTGTACCTCAAACCT
CTAATCTTCATACCCACCATCACCTCTCAAACGCTCATGCCTTCAACACCTCTTTCGAGG
CTCTGGTTCATCCAGTTCTTTTGGTAAGAAAAGAGGCCAAGATTTCAATGAAGGTTTCAG
GGAATAGAAGACATAAGCGTATGATCAAGAACAGAGAATCTGCAGCTCGTTCGCGCGCTA
GGAAACAGGCTTATACAAACAGGTTAGAACCTGAAGTTGCTCACTTGCAGGCAGAAAATG
CAAGACTCAAGAGACAACAAGATCAAAAAATGGCTGCAGCAATTCAGCAACCCAAAAAGA
ACACACTTCAACGGTCTTCCACAGCTCCATTTTGAGAAATCTACAAGTCTTGTCTCTCT
TTTGGGGATTGAGATTGTCTCATGAAGAAGTGAAAAATGGCAAAAGTTGTACCTTTT

TTATTAGCTATAAGTATAACTAAGCCTAAAATTGTAGAACTAAGATATTGTAGGGGAAAA
AAGAAGATGTAAACAAAAGACCCGAAAGAGAAAAGGATCTTTCAATTTCTTAAGGCAC
AGGAACACCTGTCTCTGGGTCCTCTCTTAATGTTCTGTCGTTTTCTATGCAAACCTTTT
TTCACCTTCTGTAACCTTATACTTGTATTCTTG

>G568 Amino Acid Sequence (domain in AA coordinates: 215-265)
MLSSAKHQHNRHLSATNKNQTLTKVSSISSSSSSSSSSSSSSSSPLPSQDSQAQKRSL
VTMEEVWNDINLASIHHLNRHSPHPQHNEPRFRGQNHNNQNPNSIFQDFLKGSNLNQEP
PTSQTTSAPNGDSTTVTVLYSSPPPPATVLSLNSGAGFEFLDNQDPLVTSNSNLHTHH
HLSNAHAFNTSFEALVPSSSFGKKRGQDSNEGSGNRRHKRMIKNRESAARSARKQAYTN
ELELEVAHLQAENARLKRQDQKMAAAIQPKNTLQRSSTAPF*

>G580 (43..747)

CCAAAAACAAAGCATTCTATGCTATTCTGTTCTGTTCTCCAATGTTGTCATCAGCAAAG
CATAATAAGATCAACAACCATAGTGCCTTTTCAATTTCTCTTCATCATCATCATTATCA
ACATCATCTCCCTAGGCCATAACAAATCTCAAGTCACCATGGAAGAAGTATGGAAAGAA
ATCAACCTTGGTTCACTTCACTACCATCGGCAACTAAACATTGGTCATGAACCAATGTTA
AAGAACCAAAACCTTAATAACTCCATCTTTCAAGATTTCCTCAACATGCCTCTGAATCAA
CCACCACCACCACCACCACCACCTTCTCTTCCACCATTGTCACTGCTCTCTATGGCTCT
CTGCCTCTTCCGCTCTGTCCTCAGCTTAAACTCCGGTGTGGATTGAGTTTT
CTTGATACCACAGAAAATCTTCTTGCTTCTAACCCTCGCTCCTTTGAGGAATCTGCAAAG
TTTGGTTGCTTGGTAAGAAAAGAGGCCAAGATTCTGATGATACTAGAGGAGACAGAAGG
TATAAGCGTATGATCAAGAACAGAGAATCTGCTGCTCGTTCAAGGGCTAGGAAGCAGGCA
TATACAAACGAACCTTGAGCTTGAAATTGCTCACTTGACAGACAGAGAATGCAAGACTCAAG
ATACAAACAGAGCAGCTGAAATAGCCGAAGCAACTCAAAACCAAGTAAAGAAAACACTA
CAACGGTCTTCCACAGCTCCATTTTGAGAAAAATCTACTATTTCTTTTTGGGGGAGTTTC
AAGTGTCTTCTATGAAGATGAGAAAAACAGAAAAAGTTGTACATTTTAGCTAAGTTAA
TTTGTGGTGGTAAGTAATGTAAAGAAAAGTGTGTGTAGAAGAAAAGTGTCTAGAAAAAG
AAAGCAACTAACTTTCTTCTTCTCTCTGCTTCTCTATCAACTCTTTTGACTTTTGTACT
TTTTTCTTCTCTACTTAACCTCTATTATTGTAATGCCAAGTCAAGTCCTTATCTAGCTA
GTACATGAGTTTCTGTTTTCACTGGTTAAGCCAT

>G580 Amino Acid Sequence (domain in AA coordinates:162-218)
MLSSAKHNKINNHSFAFSISSSSSLSTSSSLGHNKSQVTMEEVWKEINLGLSLHYHRQLNI
GHEPMLKNQNPNSIFQDFLNMPLNQPPPPPPSSSTIVTALYGSPLPLPPATVLSLNS
GVGFEFLDTTENLLASNPRSFEEAKFGCLGKKRGQSDSDTRDRLRYKRMKNRESAARS
RARKQAYTNELELEIAHLQTENARLKIQQEQLKIAEATQNVKKTLQRSSTAPF*

>G615 (197..1252)

TTTTTCTTTTCTTTCTTTTTTGTGCTGGTGTGAGAAATTGTACGCTTACTATCTCTCTCT
CTCTCTGCCAGATTCTCTCTTTTTGATGATGTGAAAGTTGTGCTTTTGTTCCTTAAGAAA
AAGGCATATTTTTAATACTTGATTCTTGGTTCTTGATTCTTGATTCTTGGTTTTTTTAG
CTTCTTAAGTTCGGTGTATGTCGTTCTCCACCAATGACTACAACGATGGTAATAACAATGG
AGTGTACCTCTCTCTCTTTACCTTCTTCACTCTCTGGCCATCAAGACATCATTCATAA
TCCCTACAACCATCAGTTAAAAGCATCTCCGGGCCATATGGTATCAGCAGTTCCTGAATC
TCTGATCGATTACATGGCGTTTAAAGTCAAATAATGTTGTGAATCAACAAGGCTTTGAGTT
TCCTGAGGTGTCAAAGGAAATCAAGAAGGTGGTGAAGAAGGACCGACATAGCAAGATTCA
AACGGCACAAGGGATTAGAGACAGGAGGGTTAGGCTTTTTATTGGGATTGCTCGCCAATT
CTTTGATCTTCAGGATATGTTGGGGTTTGATAAAGCTAGTAAAACGTTAGACTGGCTGCT
CAAGAAGTCAAGAAAAGCCATCAAAGAGGTCGTACAAGCAAAAACCTCAACAATGATGA
TGAAGATTTTGGAAACATTGGAGGCGATGTAGAACAAGAAGAGGAGAAGGAGGAGGATGA
CAATGGCGATAAGAGCTTCGTGTATGGTTTGAGCCCCGGGTACGGTGAAGAAGAAGTGGT
ATGTGAGGCCACGAAGCAGGGATAAGAAAGAAGAAGAGTGAGTTGAGAAACATCTCATC
AAAGGGGCTAGGAGCCAAAGCTAGAGGAAAAGCAAGGAGCGAACAAAGAGATGATGGC
CTATGATAATCCAGAGACTGCCTCTGATATTACACAATCTGAAATCATGGACCCATTCAA
GAGGTCTATAGTCTTCAATGAAGGAGAAGATATGACACACCTTTTCTACAAGGAACCAAT
CGAGGAGTTTGATAATCAAGAATCTATCTTAACCAATATGACTCTACCAACGAAGATGGG
TCAAAGTTACAATCAAAATATGGGATACTTATGTTGGTAGATCAGAGTTCTAGCAGCAA
CTATAATACATTTCTGCCTCAAAATTTGGATTATAGTTATGATCAAAACCTTTTTCATGA
CCAAACCTTATATGTAGTCACCGACAAAATTTCCCAAGGTTTCTTATAAATCTCGAC
AGTTTTGAAGGACTATGCATGATCAAGTTTAAACATGTAAGCCAATATAGTCCCTTATTC

CTCTGAATGTATACAAAATCTATAGTTATGTATATCTGTTCCCTTTTAAACGTATCTTTAT
TGATCTTCTGTGCCCTTGATCAAAATTGTCAATTTAAGATTCAAGTTTGTGTAATATTTTAG
CTACAACCTTTTAAGTGGTATTATTGTAACCTTTTGAACATATATTTTGAAGATGAATAA
GAACATGTTTATATAAAAA

>G615 Amino Acid Sequence (domain in AA coordinates:88-147)
MSSSTNDYNDGNNGVYPLSLYLSSLGSHQDIIHNPYNHQLKASPGHMVSAVPESLIDYM
AFKSNNVVNQGFPEVSKEIKKVVKDRHSKIQTAGQIRDRRVLFFIGIARQFFDLQD
MLGFDKASKTLDWLLKKSRKAIKEVVQAKNLNNDDEDFGNIGGDVEQEEKEEDDNGDKS
FVYGLSPGYGEEVVCEATKAGIRKKKSELRNISSKGLGAKARGKAKERTKEMMAYDNPE
TASDITQSEIMDPFKRSIVFNEGDMTHLFYKEPIEEFDNQESILTNMTLPTKMGQSYNQ
NNGILMLVDQSSSSNYNTFLPQNLDSYDQNPFDQTLVYVTDKNFPKGFL*

>G732 (73..588)
AAAAAACCAACATAAAACATAAACTCTGTCCTTTTTTGTCTTCTTGTAACCTTTTCT
TGTTAAAAATCAATGGCGTCATCTAGCAGCACATACCGGAGCTCAAGCTCTTCCGACGGT
GGTAATAATAACCCGTCGACTCCGTCGTCACCGTCGACGAACGAAAACGTAAAAGAATG
TTATCGAACAGAGAATCTGCACGTAGGTCAAGGATGCGTAAACAGAAACACGTTGATGAT
CTAACGGCTCAGATCAATCAGCTATCAAACGACAACCGTCAGATCTTGAACAGCCTCACC
GTAACATCTCAGCTTTACATGAAGATCCAAGCCGAGAATCTGTTCTCACCCTCAGATG
GAGGAGCTTAGCACCAGACTCCAATCTCTCAACGAGATCGTTGATCTTGTTCAATCCAAC
GGTGCAGGATTTGGTGTGACACGATCGACGGCTGTGGTTTGTATGATCGTACGGTTGGG
ATCGACGGATATTACGATGATATGAATATGATGAGTAATGTTAATCATTGGGGTGGTTCG
GTTTACACTAACCAACCCATTATGGCTAATGATATCAATATGTATTGATTAAATAAAATTA
ATTAATAATAATTAGATGCCCCCTTTTTTGTCTTTTTATTTTAAATTTAGCCCCATTTGGT
GTTTTTGGGTTGGTGTGATGATGTAATTATAGTACATGCATCTTTGATTGGTTGGAAGGA
TAAATATAAACTTTATATATATATTGGGGCATATATATATGAGTTGTACTTTGCATGTAT
TGGTGTGTGTTTGTATATAATTATATGATTATATATGTTTATGTTAAAAA

>G732 Amino Acid Sequence (domain in AA coordinates: 31-91)
MASSSTYRSSSSSDGGNNPNPSDSVTVDERKRRMLSNRESARRSRMRKQKHVDDLTAQ
INQLSNDNRQILNSLTVTSQLYMKIQAENSVLTAQMEELSTRLOSLNEIVDLVQSNAGAF
GVDQIDGCGFDDRTVGIDGYDDMNMSNVNHWGGSVYTNQPIMANDINMY*

>G988 (1..1338)
ATGCTTACTTCTTCAATCCTCTAGCTCCTCCTCCGAAGATGCCACCGCTACCACCACC
GAGAATCCTCCTCTTTGTGCATCGCTCCTCCTCGCCGCAACCTCCGCCTCACATCAC
CTCCGTCGTCTTCTTTTACCCTGTCGAATTTTCGTCTCCAGTCAAACCTCACCAGCCGT
CAAAACTTACTCTCAATCCTCTCCCTTAACTCTTCTCCTCACGGCGACTCCACCGAGCGA
CTTGTAACCTCTTCACTAAGCCTTGTCGCTACGAATCAACCGTCAGCAACAAGATCAG
ACGGCTGAAACGGTTGCCACGTGGACGACGAACGAATGACGATGAGTAACTCCACGGTG
TTCACGAGCAGTGTATGCAAGAAGCAGTTCTTGTTCGAACCAAGAACAACAAATCTGAC
TTCGAGTCTTGTACTATCTTTGGCTAAACCAACTAACGCCGTTTATTTCGGTTTCGGTCAT
TTAACGGCGAACCAAGCTATCCTCGACGCGACGAGACAAACGATAACGGAGCTCTACAT
ATACTTGATTTAGATATATCACAAGGACTTCAATGGCCTCCATTGATGCAAGCCCTAGCA
GAGAGGTCATCAAACCTAGCAGTCCACCTCCATCTCTCCGCATAACCGGATGCGGTCGA
GATGTAACCGGATTAAACCGAATGAGACCGGTTAACCCGGTTTCGCTGACTCTTTAGGT
CTCCAATTCCAGTTTACACGCTAGTGTATGATAGAGAAGATCTCGCCGGACTTTTGCTA
CAGATCCGATTGTTAGCTCTCTCAGCCGTACAAGGAGAGACCATTGCCGTCAATTGTGTT
CACTTCTCCACAAAATATTTAACGACGATGGAGATATGATCGGTCACTTCTTGTACGCG
ATCAAGAGCTTAACTCTAGAATCGTTACAATGGCAGAGAGAGAAGCTAATCATGGAGAT
CACTCGTTCTTGAATAGATTCTCTGAGGCAGTGGATCATTACATGGCGATCTTTGATTG
TTGGAAGCGACGTTGCCGCAATAGCCGAGAGAGACTAACCTTAGAGCAACGGTGGTTC
GGTAAGGAGATTTTGGATGTTGTGGCGGCGGAAGAGACGGAGAGAAAGCAAAGACATCGG
AGGTTTGAGATTTGGGAAGAGATGATGAAGAGGTTTGGTTTCGTTAAGCTTCTTATTGGA
AGCTTTGCTTTGCTCAAGCTAAGCTTCTTCTTAGACTTCATTATCCTTCAGAAGGTTAT
AATCTTCAGTTCTCTTAAACAAATCTTTGTTTCTTGGCTGGCAAAATCGTCCCCTCTTCTCC
GTTTCGTCGTGGAATGA

>G988 Amino Acid Sequence (domain in AA coordinates:178-195)
MLTSFKSSSSSEDAATTENPPPLCIASSAATSASHHLRRLLFTHANFVSQSNFTAA
QNLLSILSLNSSPHGDSRTERLVHLFTKALSVRINRQQDQTAETVATWTTNEMTMSNSTV

FTSSVCKEQFLFRITKNNNSDFESCYYLWLNQLTPFIRFGHLTANQAILDATETNDNGALH
 ILDLDISQGLQWPPLMQALAEERSSNPSSPPSLRITGCGRDVTGLNRTGDRLTRFADSLG
 LQFQFHTLVIVEEDLAGLLQLIRLLALS AVQGETIAVNCVHFLHKIFNDDGDMIGHFLSA
 IKSLSNRIVTMAEREANHGDSFLNRFSEAVDHYMAIFDSLEATLPPNSRERLTLEQRWF
 GKEILDVVAEETERKQRHRRFÈIWEEMMKRFGFVNVPIGSFALSQAKLLLRHLHPSEGY
 NLQFLNNSLFLGWQNRPLFSVSSWK*

>G1519 (1..1146)

ATGAGGCTTAATGGGGATTCGGGTCCGGGTCAGGATGAACCCGGTTCGAGCGGGTTTCAC
 GGCGGAATCAGACGATTCCCGTTAGCAGCTCAGCCGGAGATTATGAGAGCTGCTGAGAAA
 GACGATCAATACGCTTCTTTCATCCACGAAGCTTGCCGCGATGCCTTCCGACACCTTTTC
 GGTACAAGAATCGCTCTTGCTTACCAGAAGGAGATGAAGCTACTTGACAGATGCTTTAC
 TATGTTCTTACGACAGGTT CAGGGCAACAACTTTAGGAGAGGAATATTGTGACATTATA
 CAGGTTGCAGGGCCTTATGGACTCTCTCTACACCAGCTAGACGTGCTTTGTTTCATATTG
 TACCAGACCGCAGTTCCATATATCGCAGAGAGAATTAGCACTCGAGCTGCTACGCAAGCA
 GTCACCTTTGATGAGTCTGATGAGTTTTTTGGTGATAGTCATATCCACTCACCAAGAATG
 ATAGATCTTCCATCTTCATCTCAAGTTGAACTTCAACTTCTGTAGTATCTAGGTTAAAC
 GATAGACTTATGAGATCGTGGCACCAGCTATT CAGCGATGGCCTGTGGTTCTTCTCTGTT
 GCGCGCAAGTCTTACAACCTGGTTTTGCGTGCCAATCTGATGCTCTTCTACTTTGAAGGT
 TTTTATTATCATATATCGAAACGTGCATCCGGGGTTCGTTATGTTTTTCATAGGAAAGCAA
 CTGAATCAGAGACCTAGATACCAAATCTTGGGGTTTTCTTCTAATCCAATTGTGCATC
 CTTGCTGCTGAGGGCTTGCGTCCGAGTAATTTGTCATCTATCACTAGCTCCATT CAGCAG
 GCTTCTATAGGATCTTATCAAACTTCAGGAGGGAGAGGTTTACCTGTTTTAAATGAAGAG
 GGAATTTGATAACTTCGGAAGCTGAAAAGGGAACTGGTCTACCTCCGATTCAACTTCA
 ACGGAGGCAGTAGGGAAATGCACTCTCTGCTTAAAGCACCCGTCAGCACCCAACGGCCACT
 CCTTGTGGTCATGTGTTTTGTTGGAGCTGCATTATGGAATGGTGAACGAGAAGCAAGAA
 TGCCCTCTTTGTGCAACGCCCAATACCCATTCAAGTTTGGTTTTGTTGTATCATTCTGAT
 TTTTAG

>G1519 Amino Acid Sequence (domain in AA coordinates: 327-364)

MRLNGDSGPGQDEPGSSGFHGGIRRFPLAAQPEIMRAAEKDDQYASFIHEACRDAFRHLF
 GTRIALAYQKEMKLLGQMLYYVLTTGSGQQTLGEEYCDIIQVAGPYGLSPTPARRALFIL
 YQTAVPYIAERISTRATQAVTFDESDEFFGDSHIHSPRMLDLPSSSQVETSTSVVSRIN
 DRLMRSWHRAIQRPVVLPAVAREVLQLVLRANLMLFYFEGFYHISKRASGVRYVFIGKQ
 LNQRPRYQILGVFLLIQLCILAAEGLRRSNLSSITSSIQQASIGSYQTSGGRGLPVLNEE
 GNLITSEAEKGNWSTSDSTSEAVGKCTLCLSTRQHPTATPCGHVFCWSCIMEWCNEKQE
 CPLCRTPNTHSSLVCLYHSDF*

>G374 (1..1359)

ATGGACAACAAAATGATCAGGATATTGATGTTAGATCAGTGGTTGAAGCTGTTTCCGCC
 GATCTTTCCCTTTGGTGTCTCCCTCTATGTGGTTGAGAGCATGTGCATGCGCTGCCAAGAA
 AATGGAACAACAGATTCTTATGACCTTAATTCCTCAGTTCAGAAAGGTTCTTAATATCT
 GCATTTGAATGTCCGCATTGCGGGGAAAGGAATAATGAAGTTCAGTTCCGAGGCGAGATT
 CAACCCCGTGATGCTGTTACAATCTAGAGGTTCTAGCTGGTGATGTGAAGATATTTGAC
 CGGCAAGTTGTGAAATCTGAATCAGCCACTATTAAGATTCTGAACTGGATTTTGAGATT
 CCACCAGAGGCCCAACGTGGAAGTTTGTCTACTGTGGAAGGGATATTAGCACGGGCTGCT
 GATGAACTGAGTGCCCTTCAAGAAGAACGCAAGAAAGTTGATCCTAAACTGCTGAAGCA
 ATAGACCAATTCTTGTCCAAACCTGAGAGCTTGTCCTAAAGCAGAGACATCCTTCACCTTC
 ATTTTGGATGATCCTGTGGAACAGTTTCATTGAGAACCACATGCTCCATCACCAGAT
 CCCTCTCTAACCATCAAAATCTATGAGCGAACACCAGAGCAACAGCAACACTTGGATAT
 GTTGCTAACCCATCTCAGGCTGGACAATCAGAAGGAAGCCTTGGCGCACCTGTGATGACT
 TTCCCTTCAACTTGCGGAGCATGTACGGAGCCGTGTGAGACACGGATGTTCAAAATAGAA
 ATCCCGTACTTTTCAGGAAGTTATTGTGTCATGGCATCTACATGTGACAGTTGTGGCTATCGT
 AATTCTGAGTTGAAGCCTGGTGGTGCAATTCCTGAAAAGGGAAAGAGATTACTCTCTCT
 GTGAGGAACATTACAGACCTTAGCCGAGATGTTATCAAGTCGGACACTGCAGGAGTGATA
 ATCCCAGAACTTGATCTGGAGCTAGCTGGTGGTACACTTGGTGGAAATGGTAACAACAGTT
 GAAGGGTTGGTTACACAGATCAGAGAAAGCCTAGCGAGAGTTCACGGATTCACTTTTGGT
 GATAGTATGGAAGAGAGTAAGTTGAACAAATGGAGAGAATTTGGAGCCAGGCTCACTAAG
 CTCCTAAGCTTTGAACAGCGGTGACATTGATTCTTGATGATGAATTAGCAAATTCCTTT
 ATTGCACCAGTAACAGATGATATCAAAGATGACCATCAGCTCACATTTGAAGAGTACGAG

BNSDOCID: <WO__03013227A2_1_>

MAGFDENVAVMGEWVPRSPSPGTLFSSAIGEEKSSKRVLERELSLNHGQVIGLEEDTSSN
 HNKDSSQSNVFRGGLSERIAARAGFNAPRLNTENIRNTDFSIDSNLRSPCLTISSPGLS
 PATLLESFVFLSNPLAQSPSTTGKFPFLPGVNGNALSSSEKAKDEFDDIGASFSPHVS
 SSSSFQGTTEMMSVDYGNYNRRSSSHQSAEEVKPGSENISSNLYGIETDNQNGQNKTS
 DVTNTSLETVDHQEEEEQRRGDSMAGGAPAEADGYNWRKYGQKLVKGSEYPRSYKCTN
 PNCQVKKKVERSREGHITEIIYKGAHNHLKPPNRRSGMQVDGTEQVEQQQQQORDSAATW
 VSCNNTQQQGSNNENVEEGSTRFEYGNQSGSIQAQTGGQYESGDPVVVVVDASSTFSNDE
 DEDDRGTHGSVSLGYDGGGGGGGGEGDESESKRRKLEAFAAEMSGSTRAIREPRVVQTT
 SDVDILDDGYRWRKYGQKVVKGNPNPRSYKCTAPGCTVRKHVERASHDLKSVITTYEGK
 HNHDVPAARNSSHGCGGDSGNGNSGSSAAVSHHYHNGHHSEPPRGRFDRQVTTNNQSPFS
 RPFSPQPHLGPPSGFSGFLGQTGLVNLMPGLAYGQGKMPGLPHPYMTQPVGMSEAMMQR
 GMEPKVEPVSDSGQSVYNQIMSRLPQI*

>G1000 (1..954)

ATGGGAAGACCTCCTTGTGTGACAAGTCCAATGTCAAGAAAGGTCTCTGGACCGAGGAA
 GAAGACGCTAAGATCCTTGCTTATGTTGCTATCCATGGTGTAGGAACTGGAGCTTGATC
 CCCAAAAAGCAGGTCTGAATCGATGTGGAAAGAGCTGTAGACTAAGATGGACTAATTAC
 TTAAGACCTGACCTTAAACATGACAGCTTCTCTACCAAGAAGAAGAGCTTATCATTGAG
 TGTCATAGAGCCATTGCGCAGCAGGTGGTCTTCCATTGCACGAAAGCTTCCAGGAAGAACG
 GATAATGATGTGAAGAATCACTGGAACACAAAGCTGAAGAAGAAGCTGATGAAAATGGGG
 ATAGACCCGGTGACTCATAAACCAGTTTCTCAACTCCTTGCAGAATTCAGAAACATTAGC
 GGCCATGGAAATGCGATCCTTCAAAACAGAACCATCTAACAACCTTATACCTACACAATCC
 AACTCAGCTTGGGAAATGATGAGAAACACAACAACAACCAACCATGAGAGTTATTACACCAAC
 TCTCCAATGATGTTTACAAATTCCTCTGAGTACCAAACTACTCCATTTCAATTCTATAGC
 CATCCAATCATCTGCTCAATGGAACCATCTTTCATGCTCTTCTCATCATCTTCTACT
 AGTATCACTCAGCCAAACCAAGTACCTCAAACACCGGTTACTAATTTCTACTGGAGCGAT
 TTCCTTCTCTCGGACCCGGTTCTCAAGTAGTGGGATCCTCAGCTACTAGCGACCTCACT
 TTTACGAGAACGAACATCATTTCAACATCGAAGCCGAATACATCTCTCAAAACATCGAT
 TCAAAGGCTCGGGAACATGTCATTCCGCGAGTTCCTTCGTTGACGAAATACTAGATAAA
 GACCAAGAGATGTTGTACAGTTTCTCTCAACTCTTGAATGATTTGATTATTAG

>G1000 Amino Acid Sequence (domain in AA coordinates: 14-117)

MGRPPCCDKSNVKKGLWTEEDAKILAYVAIHGVGNWSLIPKAGLNRCGKSCRLRWTNY
 LRPDLKHDSFSTQEEELIIECHRAIGSRWSSIARKLPGRTDNDVKNHWN TKLKKLMMKG
 IDPVTHKPVSQLLAEFRNISGHGNASFKTEPSNNSILTQNSAWEMMRNTTNNHESYYTN
 SPMFTNSSEYQTTTFHFYSHPNHLLNGTTSSCSSSSSTSITQPNQVPQTPVTFYWS
 FLLSDPVPQVVGSSATS DLTFTQNEHHFNIEAEYISQNIDSKASGTCHSASSFVDEILDK
 DQEMLSQFPQLLNDFDY*

>G1067 (436..1371)

TCTCAAGCTTCTCTCTCTTTTCTCCCATAGCACATCAGAATCGCTAAATACGACTCCT
 ATGCAAGAAGAAGCTACTTCTTCTCTTGCCCTAATTAATCTACCTAAGGTTTCC
 TCTTACCTTTCATGAGAGAGATCATTTAACATAAGTCACCTTTTTTATATCTTTTGCTTC
 GTCTTTAATTTAGTCTGTTCTTGGTCTGTTTCTATATTTGTGCGCTTGCCTAACCGAT
 CACACCTTAATGCTTTAGCTATTGTTTCTCAAATCATGAGTTTTGACTTCTCGATCTG
 AGTTTTCTTTTCTCTCTTTACGCTCTTCTTACCTAGCTACCAATATATGAACGAGCAG
 GATCAAGAATCGAGAAATGATTTGAGCTGGCGAATAAGCAGTGGTGGGATAGGGAATTA
 GTAGATGCGGCGGCGATGGAAGGCGGTTACGAGCAAGGCGGTGGAGCTTCTAGATACTTC
 CATAACCTCTTTAGACCGGAGATTCAACCAACAGCTTCAACCGCAGGGCGGGATCAAT
 CTTATCGACCAAGCATCATCAGCACCAGCAACATCAACAACAACAACCAACCGTCGGAT
 GATTCAAGAGAATCTGACCATTCAAACAAGATCATCATCAACAGGGTCGACCCGATTCA
 GACCCGAATACATCAAGCTCAGCACCGGGAAAAACGTCCACGTGGACGTCCACCAGGATCT
 AAGAACAAGCCAAGCCACCGATCATAGTAACCTCGTGATAGCCCCAACGCGCTTAGATCT
 CACGTTCTTGAAGTATCTCCTGGAGCTGACATAGTTGAGAGTGTTCACGTACGCTAGG
 AGGAGAGGGAGAGCGTCTCCGTTTGTAGGAGGAAACGGCACCCTATCTAACGTCACTCTC
 CGTCAGCCAGTCACTCCTGGAAATGGCGGTGGTGTGTCGGAGGAGGAGGAGTTGTGACT
 TTACATGGAAGGTTTGTGATTTCTTCGCTAACGGGACTGTTTTGCCACCTCCTGCACCG
 CCTGGTGCCGGTGGTTTGTCTATATTTTAGCCGGAGGGCAAGGTGAGTGGTGGGAGGA
 AGCGTTGTGGCTCCCTTATTGTCATCAGCTCCGGTTATATACTAATGGCGGCTTCGTTCTCA
 AATGCGGTTTTTCGAGAGACTACCGATTGAGGAGGAGGAAGAAGAAGGTGGTGGTGGCGGA

GGAGGAGGAGGAGGAGGGCCACCGCAGATGCAACAAGCTCCATCAGCATCTCCGCCGTCT
GGAGTGACCGGTCAGGGACAGTTAGGAGGTAATGTGGGTGGTTATGGGTTTTCTGGTGAT
CCTCATTTGCTTGGATGGGGAGCTGGAACACCTTCAAGACCACCTTTTAAATTGAATTTT
AATGTCCGGAATTTATGTGTTTTTATCATCTTGAGGAGTCGTCTTTCCTTTGGGATATT
TGGTGTTTAATGTTTAGTTGATATGCATATTTT

>G1067 Amino Acid Sequence (domain in AA coordinates: 86-93)
MEGGYEQGGGASRYFHNLFRRPEIHHQQLQPQGGINLIDQH HHQHQQHQQQQPSDDSRES
DHSNKDHHQQGRPDSDPNTSSSAPGKRPRGRPPGSKNKAKPPIIVTRDSPNALRSHVLEV
SPGADIVESVSTYARRRGRGVSVLGGNGTVSNVTLRQPVTTPGNGGGVSGGGGVVTLHGRF
EILSLTGTVLPPPAPPAGGLSIFLAGGQGVVGGSVVAPLIASAPVILMAASFNAVFE
RLPIEEEEEGGGGGGGGGPPQMQQAPSASPPSGVTGQQLGGNVGGYGFSGDPHLLG
WGAGTPSRPPF*

>G1075 (19..876)
TTTGTGTTTGGTGTGGCATGGCTGGTCTCGATCTAGGCACAACCTTCTCGCTACGTCCAC
AACGTCTGATGGTGGCGGCGGCGGACAGTTCACCACCGACAACCACCACGAAGATGACGGT
GGCGCTGGAGGAAACCACCATCATCACCATCATAATCATAATCACCATCAAGGTTTAGAT
TTAATAGCTTCTAATGATAACTCTGGACTAGGCGGCGGTGGAGGAGGAGGGAGCGGTGAC
CTCGTCATGCGTCGGCCACGTGGCCGTCCAGCTGGATCGAAGAACAAACCGAAGCCGCCG
GTGATTGTACACGCGGAGAGCGCAACACTCTTAGGGCTCACATTCTTGAAGTTGGAAGT
GGCTGCGACGTTTTTCGAATGTATCTCCACTTACGCTCGGAGACAGCGCGGGATTGTC
GTTTTATCCGGGACGGGAACCGTCACTAACGTCAGCATCCGTCAGCCTACGGCGGCGCGGA
GCTGTTGTGACTCTGCGGGGTACTTTTGAGATTCTTCCCTCTCCGGATCTTTCTTCCG
CCACCTGCTCCTCCAGGGGCGACTAGCTTGACGATATTCTCGCTGGAGCTCAAGGACAG
GTCGTCGGAGGTAACGTAGTTGGTGAGTTAATGGCGGCGGGCCGGTAATGGTCATGGCA
GCGTCTTTTACAAACCTGGCTTACGAAAGGTTGCCTTTGGACGAGCATGAGGAGCACTTG
CAAAGTGGCGGCGGCGGAGGTGGAGGGAATATGTACTCGGAAGCCACTGGCGGTGGCGGA
GGGTTGCCTTTCTTTAATTTGCCGATGAGTATGCCTCAGATTGGAGTTGAAAGTTGGCAG
GGGAATCACGCCGCGCGCGGTAGGCTCCGTTTTAGCAATTAAGAACTTTAATTGTTT
TTTCCACTTTTTTGTGTTTTCTCCGAATTTATGAAATTATGATTTAAGAAAAAAACGAT
ATTGTTTCATGTATTGACCCTCTTACTGCATGGTTTTCTTCTATTGGGTTAATTGGCTAGCT
CATAAGAATTGTTTAAATTTGGTTATTGTTCATCAAATTTGCCACATATAAAGCTTCTAGC
AAAT

>G1075 Amino Acid Sequence (domain in AA coordinates: 78-85)
MAGLDLGTTSRYVHNVDGGGGGQFTTDNHHEDDGGAGGNHHHHHHNHNHHQGLDLIASND
NSLGGGGGGGSGDLVMRRPRGRPAGSKNPKPPVIVTRESANTLRAHILEVSGCDVFE
CISTYARRRQRGICVLSGTGTVINVSIRQPTAAGAVVTLRGTFEILSLSGSFLPPPAPP
ATSLTIFLAGAQGVVGGNVVGGELMAAGPVMVMAASFNTVAYERLPLDEHEEHLQSGGGG
GGGNMYSEATGGGGGLPFFNLPMSPMQIGVESWQGNHAGAGRAPP*

>G1266 (62..718)
CAATCCACTAACGATCCCTAACCGAAAACAGAGTAGTCAAGAAACAGAGTATTTTTCTA
CATGGATCCATTTTAAATTCAGTCCCCATTTCTCCGGCTTCTACCGGAATATTCTATCGG
ATCTTCTCCAGATTCTTCTCATCCTCTTCTTCTAACAATTAATCTCTTCCCTTCAACGA
GAACGACTCAGAGGAAATGTTTCTCTACGGTCTAATCGAGCAGTCCACGCAACAAACCTA
TATTGACTCGGATAGTCAAGACCTTCCGATCAAATCCGTAAGCTCAAGAAAGTCAGAGAA
GTCTTACAGAGGCGTAAGACGACGCGCATGGGGGAAATTTCGCGGCGGAGATAAGAGATT
GACTAGAAACGGTATTAGGGTTTGGCTCGGGACGTTTCAAAGCGCGGAAGAGGCGGCTTT
AGCCTACGATCAAGCTGCTTTCTCGATGAGAGGGTCTCGGCGATTCTCAATTTTTCGGC
GGAGAGAGTTCAAGAGTCGCTTTTCGAGATTAAATATACCTACGAGGATGGTTGTTCTCC
GGTTGTGGCGTTGAAGAGGAAACACTCGATGAGACGGAGAAAGACCAATAAGAAGACGAA
AGATAGTGACTTTGATCACCGCTCCGTGAAGTTAGATAATGTAGTTGTCTTTGAGGATTT
GGGAGAACAGTACCTTGAGGAGCTTTTGGGGTCTTCTGAAAATAGTGGGACTTGGTGAAA
GATTAGGATTTGTATTAGGGACCTTAAGTTTGAAGTGGTTGATTAATTTTAACCTAATA
TGTTTTTGTGTTGCTTAAATATTGATTCTATTGAGAAACATCGAAAACAGTTTGTATGT
ACTTTTGTGATACTTGGCG

>G1266 Amino Acid Sequence (domain in AA coordinates: 79-147)
MDPFLIQSPFGFSPEYSIGSSPDSFSSSSNNYSLPFNENDSEEMFLYGLIEQSTQQTY
IDSDSQDLPIKSVSSRKSEKSYRGVRRRPWGKFAAEIRDSTRNGIRVWLGTFFESABEEAL

AYDQAAFSMRGSSAILNFS AERVQESLSEIKYTYEDGCS PVVALKRKHS MRRRMTNKKTK
DSDFDHRSVKLDNVVVFEDLGEQYLEELLGSSENSGTW*

>G1311 (41..757)

AAGTATAATAACACAAAGAAACAGAGTAAAGAAAGAAAAATGGATTTTAAGAAGGAAGA
AACACTTCGTAGAGGGCCATGGCTCGAAGAAGAAGACGAACGGCTAGTGAAGGTCATTAG
TCTTTTGGGAGAACGTCGTGGGATTCTTTAGCAATAGTTTCCGGTTTGAAGAGGAGTGG
TAAGAGTTGCAGGCTAAGGTGGATGAACATCTGAATCCGACTCTGAAGCGTGGACCGAT
GAGTCAAGAAGAAGAGAGAATCATCTTTCAGCTCCATGCTCTATGGGGTAACAAGTGGTC
GAAGATTGCGAGAAGATTACCCGGTAGGACTGATAACGAGATAAAGAACTATTGGAGAAC
TCATTATAGAAAGAAACAGGAAGCTCAAACTATGGAAAGCTCTTTGAGTGGAGAGGAAA
TACAGGAGAAGAATTGTTGCACAAGTATAAGGAAACAGAGATCACTAGGACAAAGACGAC
GTCTCAAGAACATGGTTTTGTTGAAGTTGTGAGCATGGAAGTGGTAAAGAAGCCAACGG
TGGTGTGGTGGGAAGAGAAAGCTTCGGTGTATGAAATCACCGTATGAAAATCGGATTTC
GGATTGGATATCAGAGATTCTACTGACCAGAGTGAAGCAAATCTTTCAGAAGATCACAG
CAGCAATAGCTGCAGTGAGAACAATATTAACTTGGTACTTGGTGGTTTCAAGAGACTAG
GGACTTTGAGGAGTTTTTCATGTTCTCTATGGTCATAATTCTAAAGTTGGTTTATTACTT
TTTTAAAAAAAAAAAAAAAAA

>G1311 Amino Acid Sequence (domain in AA coordinates: 11-112)

MDFKKEETLRGPWLEEDERLVKVISLLGERRWDSLAIVSGLKRSKSCRLRWMNYLNP
TLKRGPMSEQEERIIFQLHALWGNKWSKIARRLPGRTDNEIKNYWRTHYRKKQEAQNYGK
LFWRGNTGEELLHKYKETEITRKTTSQEHGFVEVVSMSGKEANGGVGGRESFGVMKS
PYENRISDWISEISTDQSEANLSEDHSSNSCENNINIGTWWFQETRD FEFSCSLWS*

>G1321 (72..803)

GTTCTTGATTGGTTGGATCGGTATACCTTAGTTGATTACGTAATTAAATAGATCGGCGT
GAAGAAGAAAAATGATCATGTGCAGCCGAGGCCATTGGAGACCAGCTGAAGACGAGAAGC
TCAAGGATCTTGTGCAACAATACGGTCCTCACAATTGGAACGCCATTGCTCTCAAGCTTC
CTGGTCTGCTCTGGTAAGAGTTGTAGATTGAGATGGTTTAATCAATTGGATCCAAGGATCA
ACCGAAACCTTTTCACGGAAGAAGAAGAAGAAAGACTTTTAGCGGCTCATCGGATCCATG
GGAACAGATGGTCCATCATCGCAAGGCTTTTCCCTGGAAGAACTGATAACGCCGTCAGA
ACCATTGGCACGTCATCATGGCTCGTCGCACACGCCAAACCTCTAAGCCTCGTCTTCTTC
CCTCGACGACTTCGTCTTCTTCTTTAATGGCGAGTGAACAAATCATGATGAGTTCTGGTG
GTTATAATCATAAATTATAGTTCCGATGATCGGAAGAAAATATTTCAGCAGACTTTATAA
ATTTCCCTTACAAATTCTCTCATATCAATCATCTTCACTTCTTAAAGGAGTTTTCCTCCG
GAAAGATCGCTTTAAGTCACAAAGCAAATCAGAGTAAGAAGCCTATGGAGTTCTACAATT
TTCTACAAGTAAACACAGATTCAAACAAGAGCGAGATTATAGATCAAGATTCAAGTCAAA
GCAAACGCAGTGACTCGGACACCAAACATGAAAGTCATGTTCCATTCTTCGACTTTTAT
CCGTTGGAAGCTCTGCCTCCTAGGATTAGTTTTTTTTGCAGTAACTCCTAAATTTCTAGAT
TAACTATTTAGTCCGTATACGTACGAGATTATCTAGGTCGTTAGCATGTATGCTTGATGT
GTATAACTACTAACTAGTGAGCTATTACCTGCGAAAATTGTAAGAAAAATACATAATGTT
GATGTATCACACATTCTCAATGTCGTGTAATAATTTCCATCGAGTTGTTAACTATCAAAGTT
ATCCGTTTGAAAAAAAAAAAAA

>G1321 Amino Acid Sequence (domain in AA coordinates: 4-106)

MIMCSRGHWRPAEDEKLKDLVEQYGPHNWNIAIALKLPGRSGKSCRLRWFNQLDPRINRNP
FTEEEERLLAAHRIHGNRWSIIARLFPGRTDNAVKNHWHVIMARRTRQTSKPRLLPSTT
SSSSLMASEQIMSSGGYNHNYSSDDRKKIFPADFINFPYKFSHINHLHFLKEFFPGKIA
LSHKANQSKKPMEFYNFLQVNTDSNKSEIIDQDSGQSKRSDSKHESHVPFFDFLSVGN
SAS*

>G1326 (32..784)

CGACGGTACGGTGGATAGAGATAGCATCCATGGAGATGTCTAGAGGAAGCAACAGTTT
TGACAATAAGAAGCCTAGTTGCCAAAGAGGTCACTGGAGACCTGTTGAAGATGACAATCT
CCGGCAACTCGTTGAACAATACGGTCCCAAGAACTGGAATTTTATTGCTCAACATCTCTA
TGGAAGATCAGGGAAGCTGTAGATTAAGATGGTACAACCAACTTGATCCAAACATCAC
CAAGAAACCCTTCACCGAGGAGGAAGAAGAGAGACTGCTTAAAGCTCATCGGATCCAAGG
GAATCGTTGGGCCTCCATAGCCCGACTGTTCCCGGGAGGACCGACAACGCTGTCAAAAA
CCATTTTCATGTCATCATGGCTAGACGCAACCGGGAAGAACTTCTCTCCACAGCTACTTC
TACGTTCAACCAAACTTGGCATACTGTTTGGAGCCCTAGTTCTAGTCTTACAAGGCTAAA
TAGATCCCATTTGGGCTATGGAGGTATCGAAAGGATAAGAGTTGCGGTCTCTGGCCCTTA

CTCTTTTGTTCACCACTACGAATGGTCAATTTGGATCTTCATCTGTCTCTAACGTACA
 CCACGAAATTTATCTTGAGAGGAGAAAGTCGAAAGAGTTGGTGGATCCTCAGAATTACAC
 ATTTTCATGCAGCCACACCAGATCATAAGATGACTTCAAATGAAGATGGACCATCCATGGG
 AGATGATGGTGAGAAGAACCAGTGTTACTTTTCATTGATTTTCTTGGTGTGGATTAGCTTC
 TTAGGTTATAACATCACAAGTCAAAGCTTTTAAGGGTTTCTATCATTAGGGTTAGGCATC
 ATTTTCAGCCTTTTGCTTCCTTAAACTCTCATATGGATCT

>G1326 Amino Acid Sequence (domain in AA coordinates: 18-121)
 MEMSRGNSFDNKKPSCQRGHWRPVEDDNLRLQVLEQYGPKNWNFIAQHLYGRSGKSCRLR
 WYNQLDPNITKKPFTEEEERLLKAHRIQGNRWASIALRFPGRTDNAVKNHFHVMARRK
 RENFSSSTATSTFNQTWHTVLSPPSSSLTRLNRSHFGLWRYRKDKSCGLWPYSFVSPTNGQ
 FGSSSVSNVHHEIYLERRKSKELVDPQNYTFHAATPDHKMTSNEDGPSMGDDGEKNDVTF
 IDFLGVGLAS*

>G1367 (128..1567)
 TCCTTCCACAAAATTTTTTAAATTTTATCTGAAAAATTAAAAACAACCGAAACAAAAAAA
 AAAACTAAAAATCAAAAATCTCATCACCTTCCTTGCTCTGTATTTTTTCTCTCTCACTAA
 ATCCTCCATGGATCCTTCTCTCTCTGCAACCAATGATCCTCATCATCCTCCTCCTCCTCA
 GTTCACATCTTTCCCTCCTTTTACCAACACCAACCCCTTCGCCTCTCCAAACCACCCCTT
 CTTACCCGACCCACCGCCGTCGCGCCGCCAAACAACATCCATCTCTATCAAGCAGCTCC
 TCCGACAGCCACAAACATCTCCAGTTCCTCCTCATCCATCTATTTCCACCCCTCCTTA
 CTCTGACATGATTTGCACGGCGATTGCAGCGTTAAACGAACAGATGGGTCAAGCAAGCA
 AGCTATTTTCAGAGTACATAGAGAGAATTTTACACTGGGATTCTCTACTGCTCATGGAGCTTT
 GTTGACACACCATCTCAAGACTTTGAAGACCAGTGGGATTCTTGTCTATGGTTAAGAAATC
 TTACAAGCTTGCTTCTACTCCTCCTCCTCCTCCTACTAGTGTAGCTCCTAGTCTTGA
 ACCTCCAGATCTGATTTTCATAGTCAACGAGAACCAACCTTTACCTGATCCGGTTTTGGC
 TTCTTCTACTCCTCAGACTATTAACCGTGGTTCGTGGTTCGACCTCCAAAAGCTAAACCAGA
 TGTGTTTCAACCTCAACCTCTGACTAATGGAAGAACTCACCTGGGAACAGAGTGAATTACC
 TGTCTCTCGACAGAGGAGATACAGATACAGCCGCCACAGTTACCGTTACAGCCACAGCA
 GCCGGTTAAGAGACCCGCCGGTTCCTTAGAAAAGATGGAACCTTCGCCGACGGTGAAGCC
 AGCTGCTTCTGTTTTCCGGTGGTGTGGAGACTGTGAAACGAAGAGGTAGACCTCCGAGTGG
 AAGAGCTGCTGGGAGGGAGAGAAAAGCCTATAGTAGTCTCAGCTCCAGCTTCAGTGTTCCT
 GTATGTTGCTAATGGTGGTGTAGACGCCGAGGGAGACCAAAGAGAGTTGACGCTGGTGG
 TGCTTCTCTGTTGCTCCACCACCACCACCACCACTAACGTAGAGAGTGGAGGAGAGGA
 GGTTGCAGTCAAGAAACGAGGAAGAGGACGGCCTCCTAAGATTGGAGGTGTTATCAGGAA
 GCCTATGAAGCCGATGAGAAGCTTTGCTCGTACTGGAAGAACCCGTAGGAAGACCCAGAAA
 GAATGCGGTGTCTAGTGGGAGCTTCTGGACGACAAGATGGTGAATGAGAGAACTGAAGAA
 GAAGTTTGAAGTTGTTTCAAGCGAGAGCTAAGGATATTGTAATTGTGTTGAAATCCGAGAT
 AGGAGGAAGTGGAAATCAAGCAGTGGTTCAAGCCATACAGGACCTGGAAGGGATAGCAGA
 GACAACAAACGAGCCAAAGCACATGGAAGAAGTGCAGCTGCCAGACGAGGAACACCTTGA
 AACCGAACCAAGCAGAGGGTCAAGGACAGACAGAAGCAGAGGCAATGCAAGAAGCTCT
 GTTCTAAAGATAAAGCCTTGACATAAAAAGCTAGCAAGTGGTGGGTTTACTTGTGTGTG
 TTACATGAAATTTTTAATCTTATAAGGGTGTGTTGCAGGAGAAAAACAAAAGAACAATGT
 GATGAACTGATGATGATGATTGTGTCTCTAACCACCAACAAGGAGAGGTAGGGTAATGT
 CTGTAAAGTGAATTAGGATGTTACCATGTTTCATGCTTCCCATCTCTCCTCATCGTCCAT
 ATCTGTGTAGGCAGCTTTGTTCTTGTTCCTCGTGTGTTTTTTTAGACTGTTGTGTCTCT
 TATTCTATTTGTCTCCTTAGGCTTTTTTAGGAGTTGTTGTTGATGTTTATCAAAAACGCT
 TATGTAATTTTTATGACCACTTCTACTTTTTATGATGGTTTTCTT

>G1367 Amino Acid Sequence (domain in AA coordinates: 179-201, 262-285, 298-319, 335-357)

MDPSLSATNDPHHPPPPQFTSFPPFTNTNPFASPNHPFFTGPTAVAPPNNIHLYQAAPPQ
 QPQTSPPVPHPSISHPPYSMDICTAIAALNEPDGSSKQAIISRYIERIYTGIPAHGALLT
 HHLKTLKTSGLIVMVKKSYYLASTPPPPPTSVAPSLEPPRSDFIVNENQPLPDPVLASS
 TPQTIKRGRRPPKAKPDVQPPQLTNGKLTWEQSELVSRPEEIQIOPPQLPQPQPV
 KRPPGRPRKDGTSPTVKPAASVSGGVETVKRRGRPPSGRAAGRERKPIVVSAPASVFPYV
 ANGGVRRRRGRPKRVDAGGASSVAPPPPPPTNVESGGEEVAVKKRGRGRPPKIGGVIRKPM
 KPMRSFARTGKPVGRPRKNAVSVGASGRQDGDYELKKKFELFQARAKDIVIVLKSEIGG
 SGNQAVVQAIQDLEGIAETTNEPKHMEEVQLPDEEHLETEPEAEGQGQTEAEAMQEALF*
 >G1386 (89..673)

AATTTTATTTCTCTCTCAAATCTTCCCACCAAAAATTAACCTTTTCGTTTCACTAAG
 TCCCTTTTAAAAGAAAATATCCCAATTAATGGAACGTGACGACTGCCGGAGATTTTCAGGA
 CTCGCCGGCGCAGACGACGAGAGAGAAGAGTGAAATATAAACCAAGAAGAAAAGAGCCAA
 AGATGATGATGATGAGAAAAGTTGTTTCGAAGCATCCAAATTTTCGAGGTGTCAGAATGAG
 ACAATGGGGAAAATGGGTGTCCGAAATCAGAGAGCCAAAAAGAAATCAAGAATCTGGCT
 CGGTACTTTCTCCACGGCGGAGATGGCGGCGGTGCTCACGACGTGGCAGCTTTAGCCAT
 CAAAGGCGGTTCTGCACATCTCAACTTCCCGAGCTCGCTTATCACCTCCCTAGACCAGC
 TAGTGCCGACCCTAAAGACATCCAAGCTGCCGCCGCGCAGCTGCAGCCGCTGTGGCCAT
 TGACATGGATGTAGAGACGCTTTCGCCGTGCCATCTCCACAGTTACGGAAACGTCATC
 TCCGGCTATGATAGCACTCTCCGACGACGCGTTCTCCGATCTTCCTGATCTCTTGCTCAA
 CGTGAACCATAACATCGATGGCTTCTGGGACTCTTTTCCCTATGAAGAACCTTCTCTC
 TCAAAGTTACTAGAAATCAAACTATGTCGTTTTTGTATGTATTTTGTATGTGACCA
 TTTTGTGACGTGCAAAATCACCCGGATAATCCAAATTGTATGATTTATTAATGGTTGATG
 ATTTCTTTGTGTGGAACAATGTGTATGATACGTAATCAAAAGTTCAAAAAAAAATAAA
 AAAAA

>G1386 Amino Acid Sequence (domain in AA coordinates: TBD)
 MERDDCRRFQDSPAQTERRVKYKPKKKRAKDDDEKVVSKHPNFRGVRMRQWGKVVSEI
 REPKKKSRIWLGTFSTAEMAARAHDAALAIKGGSAHLNFPDELAYHLRPPASADPKDIQA
 AAAAAAAVAIDMDVETSSPSPPTVTETSSPAMIALSDDAFSDLPDLLLNVNHNIDGFW
 DSFPYEEFPLSQSY*

>G1421 (292..1155)
 GAAATTTTCATCCCTAAATAAGAAAAAGCATCTCCTTCTTTAGTGTCTCTCTTACCAAA
 CTCTTGATTCCATAAGCATATATTAAGCTCTCTGCTTTCTTCAACTTTCCCGGGAA
 AATCTTCTTGTACAAAGCATCAATCTCTTGTGTTTACCAATTTCTCTTATTCTCTTT
 TTTGCCCTTTACTTTTCTTAACCTTTGGTCTTTATATATAAACACACGACACAAAGAAGAA
 CACACATAAGTTAAACTATTACAACAGTTTAAAGAGAGAGATTAAAAAATGGAGACA
 GAGAAGAAAGTTTCTCTCCCAAGAAATCTTACGAATCTCTGTTACTGATCCTTACGCAACA
 GATTCTGTAAGCGACGAAGAAGAAGATTGATTTGATGCATTATCTACAAAACGACGT
 CGTGTTAAGAAGTACGTGAAGGAAGTGGTGTGATTCGGTGGTTTCTGATAAAGAGAAG
 CCGATGAAGAAGAAGAGAAAGAGCGCGTTGTTACTGTTCCAGTGGTTGTTACGACGGCG
 ACGAGGAAGTTTCGTGGAGTGAGGCAAAGACCGTGGGGAAAATGGGCGGCGGAGATTAGA
 GATCCGAGTAGACGTGTAGGGTTTGGTTAGGTACTTTTGACACGGCGGAGGAAGCTGCC
 ATTTGTTTACGATAACGCAGCTATTACGTACGTGGTCTTAACGCAGAGCTTAACCTCCCT
 CCTCCTCCCGTGACGGAGAATGTTGAAGAAGCTTCGACGGAGGTGAAAGGAGTTTCGGAT
 TTTATCATTTGGCGGTGGAGAATGCTTTCGTTCCGCCGTTTCTGTTCTCGAATCTCCGTTT
 TCCGGCGAGTCTACTGCGGTTAAAGAGGAGTTTGTGCGGTGTATCGACGGCGGAGATTGTG
 GTTAAAAAGGAGCCGTCTTTTACGGTTTCAAGATTTCTCGGCGCGGTTGTTCTCGGACGAC
 GACGTTTTTGGTTTCTCGACGTGATGAGTGAAAGTTTCGGCGGCGATTATTTGGAGAT
 AATCTTTTTCGGATATGAGTTTGGATCCGGGTTTGGATTCCGGTCTGGGTCTGGATTC
 TCCAGCTGGCAGCTTGAGGACCATTTTCAAGATATTGGGGATTATTCGGGTCCGATCCT
 GTCTTAACTGTTTAAAGAAATAACTGGCCGTTTAAACGGCGTTTAGTGAAGTTTGTACCG
 GCGACGGCGAGGATTAAAAAAAACGGCGATTATTTTGAATGAAGATTGTTAAATA
 >G1421 Amino Acid Sequence (domain in AA coordinates: 74-151)
 METEKKVSLPRILRISVTDPYATDSSSDEEEVDFDALSTKRRRVKKYVKEVVLDSVSD
 KEKPMKKKRRKRVVTVPVVTTATRKFRGVRQRPWKWAAEIRDPSRRVRVWLGTFDTAE
 EAAIVYDAAAIQLRGPNAELNFPPTVTENVEEASTEVKGVSDFIIGGGECLRSPVSVLE
 SPFSGESTAVKEEFVGVSTAEIVVKKEPSFNGSDFSAPLFSDDVDFGFSTSMSESFGGDL
 FGDNLFADMSFGSGFGFGSGSGFSSWHVEDHFQDIGDLFGSDPVLTV*

>G1453 (39..917)
 CGTCGACGCGAAATCCTAGAAAATAACTATCAATATGATGAAGGTTGATCAAGATT
 ATTCGTGTAGTATACCGCTGGATTTAGGTTTCATCCGACAGATGAAGAAGTTGTGCGGAT
 ATTATCTCAAGAAGAAAATCGCCTCCAGAGGATTGATCTCGACGTTATCAGAGAAATTG
 ATCTTTACAAGATCGAACCATGGGATCTACAAGAGAGATGTAGGATAGGGTACGAGGAGC
 AAACGGAGTGGTATTTCTTCAGCCATAGAGACAAGAAGTATCCGACTGGGACTAGGACAA
 ACCGAGCCACCGTGGCCGTTTCTGGAAAGCAACGGGCCGGGACAAGGCGGTTTACCTCA
 ACTCCAACTTATCCGTATGAGAAAAACGCTTGTCTTTTACCGAGGTGAGCGCCTAATG
 GCCAAAAGTCCGATTGGATCATTCACGAATACTACAGCCTCGAGTCACACCAGAACTCTC

CTCCACAGGAAGAAGGATGGGTAGTGTGTAGAGCATTTAAGAAACGAACGACCATCCCAA
CAAAAAGGAGGCAACTTTGGGATCCGAACGCTTATTCTACGACGACGCCACTCTCTTGG
AACCTCTCGACAAGCGAGCCAGACATAATCCTGATTTTACCGCCACACCGTTCAAGCAAG
AACTACTCTCCGAGGCCAGTCACGTCCAGGATGGAGATTTTCGGATCTATGTACCTTCAAT
GCATCGATGATGATCAATTCTCCAGCTTCTCAGCTCGAGAGCCCCCTCTCTTCCGTCGG
AAATAACTCCCCATAGTACTACTTTTTCTGAGAACAGTAGCCGGAAGATGACATGAGCT
CCGAGAAGAGGATCACTGACTGGAGATATCTAGATAAGTTTCGTGGCGTCTCAATTTTGA
TGAGTGGAGAAGACTAAAAAAGGCTTTCCTATGCATGCATGCACTAGAAACGTCGTCGCA
TTTTGGATTACATGCGGCCGCT

>G1453 Amino Acid Sequence (conserved domain in AA coordinates:13-160)

MMKVDQDYSCSIPPGFRFHPTDEELVGYYLKKKIASQRIDLDVIREIDLYKIEPWLQER
CRIGYEEQTEWYFFSHRDKKYPGTGRTNRATVAGFWKATGRDKAVYLNLSKLIGMRKTLVF
YRGRAPNGQKSDWIIHEYYSLESHQNSPPQEKGWVVCRAFKKRTTIPTKRRQLWDPNCLF
YDDATLLEPLDKRARHNPFTATPFKQELLSEASHVQDGFSGMYLQCIDDDQFSQLPQL
ESPSPSEITPHSTTFSENSSRKDDMSSEKRITDWRYLDKFVASQFLMSGED*

>G1560 (120..1340)

ATCCTTTCAATTTCCACTCCTCTCTAATATAATTCACATTTTCCCACTATTGCTGATTCA
TTTTTTTTTGTGAATTATTTCAAACCCACATAAAAAATCTTTGTTTAAATTTAAAACCA
TGGATCCTTCATTTAGGTTTCATTAAGAGGAGTTTCTGCTGGATTTCAGTGATTCTCCAT
CACCACCATCTTCTTCTTACATCTTTATTCATCTTCCATGGCTGAAGCAGCCATAAATG
ATCCAACAACATTGAGCTATCCACAACCATAGAAGGTCTCCATGAATCAGGGCCACCTC
CAATTTTGCACAAAGACATATGACTTGGTGGAAGATTCAAGAACCAATCATGTCTGTCTT
GGAGCAAATCCAATAACAGCTTCAATTGTCTGGGATCCACAGGCCTTTTCTGTAACCTCTCC
TTCCAGATTCTTCAAGCACAATAACTTCTCCAGTTTGTCCGCCAGCTCAACACATATG
GTTTCAGAAAGGTGAATCCGGATCGGTGGGAGTTTGCAAACGAAGGGTTTCTTAGAGGGC
AAAAGCATCTCCTCAAGAACATAAGGAGAAGAAAAACAAGTAATAATAGTAATCAAATGC
AACAACCTCAAAGTTCTGAACAACAATCTCTAGACAATTTTTGCATAGAAGTGGGTAGGT
ACGGTCTAGATGGAGAGATGGACAGCCTAAGGCGAGACAAGCAAGTGTGATGATGGAGC
TAGTGAGACTAAGACAGCAACAACAAAGCACCAAAATGTATCTCACATTGATTGAAGAGA
AGCTCAAGAAGACCGAGTCAAAAACAAAAACAATGATGAGCTTCTTGCCTCGCGCAATGC
AGAATCCAGATTTTATTTCAGCAGCTAGTAGAGCAGAAGGAAAAGAGGAAAGAGATCGAAG
AGGCGATCAGCAAGAAGAGACAAGACCGATCGATCAAGGAAAAGAAATGTGGAAGATT
ATGGTGATGAAAGTGTTATGGGAATGATGTTGCAGCCTCATCCTCAGCATTGATTGGTA
TGAGTCAGGAATATACATATGGAAACATGTCTGAATTCGAGATGTCCGAGTTGGACAAAC
TTGCTATGCACATTCAAGGACTTGAGAGATAATTCAGTGCTAGGGAAGAAGTCTTGAATG
TGGAAAAAGGAAATGATGAGGAAGAAGTAGAAGATCAACAACAAGGTACCATAAAGGAGA
ACAATTGAGATTATGTGAAGTTTGTGGGAAGATTGTTAAATGAAGGTCAAAATTTTG
ATTTTGAAGGAGATCAAGAAAATGTTGATGTGTTAATTCAGCAACTTGGTTATTGGGTT
CTAGTTCACACACTAATTAAGAAGAAATGAAATGATGACTACTTTAAGCATTGGAATCA
ACTTGTTCCTATTAGTAATTTGGCTTTGTTTCAATCAAGTGAGTCGTGGACTAACTTGC
>G1560 Amino Acid Sequence (domain in AA coordinates: 62-151)

MDPSFRFIKEFFPAGFSDSPSPSSSSYLYSSSMAEAAINDPTTSLYPQPLEGLHESGPP
PFLT KTYDLVEDSRTNHVSVWSKSNNSFIVWDPQAFSVTL LPRFFKHNNFSSFVRQLNTY
GFRKVNPD RWEFANEGFLRGQKHL LKNIRRRKTSNNSNQMQPQSSEQQSLDNFCIEVGR
YGLGEMDSLRRDKQVLM MELVRLRQQQSTKMYLTLIEEKLKKTESKQKQMMSFLARAM
QNPDFIQQLVEQKEKRKEIEEAI SKKRQRPIDQGRNVEDYGDSEGYGNDVAASSSALIG
MSQEYTYGNMSEFEMSELDK LAMHIQGLGDNSSAREEVLNVEKGNDEEEVEDQQQGYHKE
NNEIYGE GFWE D L L N E G Q N F D F E G D Q E N V D V L I Q Q L G Y L G S S S H T N *

>G1594 (1..984)

ATGGATGGAATGTACAATTTCCATTTCGCCCGGTGATTATTCAGATAAGTCGGTTCTGATG
ATGTCACCGGAGAGTCTCATGTTTCTTCCGATTACCAAGCTTTGCTATGTTCTCCGCC
GGTGAATTCGTGTCTCTGATGTTTTCGGATCCGACGAGCTACTCTCAGTAGCCGTCTCC
GCTTTGTCGTCCGAGGCGGCTTCGATCGCTCCGAGATCCGAAGAAATGATGATAACGTT
TCTTAACGTTCATCAAGCTAAAAATCGCTTGTCATCCTTCGTATCCTCGCTTACTTCAA
GCTTACATCGATTGCCAAAAGGTCCGAGCACCACCGAGATAGCGTGTCTTACTAGAGGAG
ATTCAACGGGAGAGTGATGTTTATAAGCAAGAGGTTGTTCTTCTTCTGCTTTGGAGCT
GATCCTGAGCTTGATGAATTTATGGAACGTACTGCGATATATTAGTGAAATACAAATCG

GATCTAGCAAGACCGTTTGACGAGGCAACGTGTTTCTTGAACAAGATTGAGATGCAGCTA
CGGAACCTATGTACTGGTGTGAGTCTGCCAGGGGAGTTTCTGAGGATGGTGTAAATATCA
TCTGACGAGGAACCTGAGTGGAGGTGATCATGAGGTAGCAGAGATGGGAGACAAAGATGT
GAAGACCGGGACCTCAAAGATAGGTTGCTACGCAAATTTGGAAGCCGTATTAGTACTTTA
AAGCTTGAGTTCTCAAAGAAGAAGAAGAAAGGAAAGTTACCAAGAGAAGCAAGACAAGCT
CTTCTTGATTGGTGAATCTCCATTATAAGTGGCCTTACCCTACTGAAGGAGATAAGATA
GCATTAGCTGATGCAACGGGGTTAGACCACAAAACAAATCAACAATTGGTTTATAAACCAA
AGGAAACGTCATTGGAAGCCATCAGAGAATATGCCTTTTCGCTATGATGGATGATTCTAGT
GGATCATTCTTTACCGAGGAATGA

>G1594 Amino Acid Sequence (conserved domain in AA coordinates:343-308)

MDGMYNFHSAGDYSDKSVLMMSPELMPSPDYQALLCSSAGENRVSDVFGSDELLSVAVS
ALSSEAASIAPEIRNDDNVSLTVIKAKIACHPSYPRLLQAYIDCQKVGAPPEIACLLLEE
IQRESDVYKQEVVPSSCFGADPELDEFMETYCDILVKYKSDLARPFDEATCFLNKIEML
RNLCTGVESARGVSEDGVISSDEELSGGDHEVAEDGRQRCEDRDLKDRLLRKFGSRISTL
KLEFSKKKKKGLPREARQALLDWWNLHYKWPYPTEGDKIALADATGLDQKQINNWFINQ
RKRHWKPSENMPFAMDDSSGSFFTEE*

>G1750 (94..1101)

CCCTTTTCTCTCTTTCTCCAAATCTCTGAAAATTTTACCAGAATCTCTGTTCTTTTTT
TCACCAGAATCTCTCTGTTTAAATAATAGGTGATGATGATGGATGAGTTTATGGATCTT
AGACCAGTGAAGTACACAGAGCACAAGACTGTTATCAGAAAGTACACTAAAAAGTCGTCT
ATGGAGAGGAAGACAGTGTTCGTGACTCGGCCAGGTTGGTTCCGGTCTCAATGACGGAT
CGTGACGCCACTGATTCATCAAGCGACGAGGAAGAGTTTCTGTTCCCTCGAAGACGTGTC
AAGAGATTGATTAACGAGATCAGAGTCGAGCCTAGCAGCTCTTCCACCGGCGACGTCTCT
GCTTCTCCGACGAAGGACCGGAAAAGAATCAACGTTGATTCTACGGTTCAAAGCCCTCT
GTTTCCGGCCAAAACCAAGAGAAGTACCGCGCGGTGAGACAGCGACCATGGGGAAAATGG
GCGGCGGAGATTCTGTATCTGAGCAACGCCGAGAATCTGGCTCGGTACTTTTGCAACG
GCGGAGGAAGCTGCCATCGTCTACGACAACGCAGCAATCAAATTCGTGGCCCTGATGCT
CTTACCAACTTCACCGTACAACCAAGAACCGGTACAAGAACAAGAACAAGAACCG
GAGAGCAACATGTCGGTTTCGATATCAGAATCAATGGACGATTCTCAACATCTATCATCT
CCGACATCGGTTCTCAACTACCAACATATGTCTCGGAGGAACCAATCGATAGTCTTATC
AAACCGGTTAAACAAGAGTTTCTTGAACCAGAACAGAGCCAATAAGCTGGCATCTTGGA
GAAGGTAATACTAATACTAATGATGATTCAATTTCCATTGGACATTACATTTCTCGACAAC
TATTTCAATGAATCATTACCAACATCTCCATCTTCGATCAACCTATGTCTCCTATTCAA
CCAAACAGAGAATGATTCTTCAACGACCTTATGTTATTTCGATAGCAACGCAGAAGAATAC
TACTCTCCGAGATCAAAGAGATTGGTTTCATCGTTCAACGATCTTGATGATTCTTTGATA
TCCGATCTCTTACTGTGTGATATTTTGGCATTAAACCAACACCGGTTTGGTTGC

>G1750 Amino Acid Sequence (domain in AA coordinates: 107-173)

MMMDDEFMDLRPVKYTEHKTVIRKYTKKSSMERKTSVRDSARLVRVSMTRDATDSSSDEE
EFLFPRRRVKRLINEIRVEPSSSSTGDVSASPTKDRKRINVDSTVQKPSVSGQNQKKYRG
VRQRPWGKWAEEIRDPEQRRRIWLGTFAAEAAIIVYDNAIKLRGPDALTNFTVQPEPE
PVQEQEQEPESNMSVSISESMDDSQLHSSPTSVLNYQTYVSEEPIDSLIKPVKQEFLEPE
QEPISWHLGEGNTNTNDDSFPLDITFLDNYFNESLPDISIFDQPMSPIQPTENDFFNDLM
LFDNSAEEYYSSSEIKEIGSSFNDLDDSLISDLLLV*

>G1947 (70..918)

ACAACTATTCTCTCTCTCTCTTTTTTTTATTAAAAAGCTCAAATTTATATAGGTTTTTT
GTTCAAAAATGGATTATAACCTTCCAATTCATTAGAGGTCTCAAAGAAACGCCACCA
ACGGCTTTCTTGACGAAAACATACAACATAGTGGAGGATTCAAGCACAAACAACATAGTT
TCATGGAGCAGAGACAAACAGCTTCATTGTTTGGGAACCAGAGACTTTTGCCCTAATT
TGCCCTCCCTAGATGCTTTAAGCACAAATAATTTCTCCAGCTTTGTTAGACAGCTCAATACT
TATGGGTTTAAAGAGATTGATACAGAGAGATGGGAATTTGCAAATGAGCATTTTCTGAAG
GGAGAGAGGCATCTTCTTAAGAACATCAAGAGAAGAAAGACATCTCAAACGCAACAG
CAGTCGCTAGAAGGAGAGATCCATGAGCTGCGAAGAGACAGAATGGCTTTAGAAGTAGAA
CTGGTTAGACTGCGACGAAAACAAGAAAGCGTGAAGACATATCTGCATTTGATGGAAGAG
AAACTGAAAGTCACAGAAGTAAAGCAAGAAATGATGATGAATTTCTTGCTAAAGAAGATT
AAGAAACCGAGTTTTTTTACAGAGCTTAAGGAAACGTAATCTGCAAGGAATCAAGAATCGA
GAGCAAAAGCAAGAGGTGATCTCAAGCCATGGTGTGAGGATAATGAAAAGTTTGTTAAA
GCTGAGCCAGAAGAGTATGGTGATGACATCGATGATCAATGTGGAGGTGTGTTTGATTAT

GGTGATGAGCTTCACATAGCTTCAATGGAGCATCAAGGACAAGGGGAGGATGAAATTGAA
ATGGATAGTGAAGGAATTTGGAAGGGTTTCGTGTGAGTGAGGAGGAGATGTGTGATTGA
GTGGAACATTTTATATAATAAACTAATGTATTATGAGAGGTTTTTTTTGTTTTTTTGCT
TTTTTTTTCCGAGTTTGTCAATGATCAAGCATTGTATACAATTTGGGCCAACTAAAAGCCCAA
CAAAATATTTGGCCTTGGCATTGTGTAACAAATTGACTAATTCGGCCACACCTTCC
>G1947 Amino Acid Sequence (domain in AA coordinates: 37-120)
MDYNLPIPLEGLKETPPTAFLTKTYNIVEDSSTNNIVSWSRDNNFIVWEPETFALICLP
RCFKHNNFSSFVRQLNTYGFKKIDTERWEFANEHFLKGERHLLKNIKRRKTSSQTQTQSL
EGEIHLLRRDRMALELVRLRRKQESVKTYLHLMEEKLVTEVKQEMMNFLKKIKKP
SFLQSLRKRNLQGIKNREKQEVISSHGVEDNGKFVKAEPPEYDGDIDQCGGVFDYGD
LHIASMEHQGGEGEIEMDSEGIWKGFVLSSEEMCDLVEHFI*
>G2011 (309..1547)
AATGTCGGTTGTACAATATTTGTCTACTAAAGTTTCCAAATTTCTTCTAAACTGATGAAT
CAATGGAACATGATGACGAAAAAGATAAATCCACGGTGGCGGGAAGTACCCACCCATTT
CCACCGCCTCTCTATTCCCCAGATTTTTTTCAATTATCTGACTACAGTTTGTCTGTTACT
TCCTTCCCTAAACCTTTATAAACCATTAACCTCTCATCTTCTTCTTAAACCCCTTA
ATTATCACACACCCCAATTTCTCACTCTCTCTCTCACTAAACCCGTAATTTTCTAC
TATATCAAATGAGCCAAAAAGATGCTGTTTCTAAACCAACTCCAATTTTCAGTACCCG
TTTCGAGACGATCCGATATACCCGGTCTCTCTACGTCGACACTGACATGGGTTTCTCTG
GGTCAACACTTCCCATGCCACTAGACATCTTACAAGGGAATCCAATTTCCACTTTTTTAT
CCAAGACTTTTGATTGTTGGTTGATGACCCGACTCTTGACCCGGTCATCTCTTGGGGACTGA
CCGGAGCTAGCTTCGTAGTTTGGGATCCTCTAGAGTTTGGCAGAATCATACTTCCAAGGA
ATTTCAAACACAACATTTCTCCAGCTTCGTCAGACAGCTTAACACTTATGGATTTGAA
AGATTGATACTGACAAGTGGGAATTCGCTAACGAGGCTTTCTTAGAGGCAAGAAGCATC
TTCTGAAGAACATTATCGTCGTCGATCACCACAATCCAACCAAACTTGCTGCAGTAGCA
CTAGCCAAAGCCAAGGGTCACCTACTGAGGTTGGAGGAGAGATTGAGAAGCTGAGGAAAG
AGCGGCGTGCAATGATGGAGGAAATGTTGAGCTTCAGCAGCAAAGCAGAGGCACAGCTC
GACATGTGGACACTGTAAACCAGAGGCTGAAAGCTGCAGAGCAACGTCAGAAAGCAATTGC
TCTCTTTCTTGGCTAAGTTGTTTCAAGAACCGGGTTTCTTGGAACGCCTGAAGAACTTCA
AAGGAAAAGAAAAGGAGGAGCTTCTGGATTGGAAGGCGAGAAAGAGTTTCATCAAGC
ACCACCAGCAGCTCAAGATTCTCAACAGGAGGGGAGGTGGTGAAGTATGAAGCTGATG
ATTGGGAGAGATTGCTAATGTATGACGAAGAGACTGAGAACACCAAGGGTTTAGGAGGGA
TGACTTCAAGCGATCCAAAAGGCAAGAACTTGATGTATCCATCAGAAGAAGAGATGAGCA
AACCAGATTACTTGATGTCTTCCCATCTCTGAAGGACTTATTAACAAGAAGAGACGA
CATGGAGCATGGGTTTTCGATACATAACCGAGTTTCAGCAACACCGATGCATGGGGAA
ACACAATGGACTATAATGATGTCTCAGAGTTTGGTTTGTCTGCAGAAACAACAAGTGATG
GTTTGCCTGATGTCTGCTGGGAACAATTTGCTGCAGGAATCAAGAGACTGGATTCAACT
GGCCAACCTGGTGATGATGATGATAATACGCCAATGAATGATCCTTAGGATCTTTTCATAT
ATAGTTTACACAAAACCCGTTCTTATCGGGTGAATATTAATTCATTATTCTTTTG
AATGCACTCTTTATACATATATATAATATTGATGAGTTTGATTGTTCCAAAAA
>G2011 Amino Acid Sequence (domain in AA coordinates: 56-147)
MSPKKDAVSKPTPIVPSVRRSDIPGSLYVD TDMGFSGLPMPDLILQGNPIPPFLSKT
FDLVDDPTLDPVISWGLTGASFVWDPLEFARIILPRNFKHNNFSSFVRQLNTYGFRIKID
TDKWEFANEAFRLRGKKHLLKNIHRRRSPQSNQTCSSSTSQSQSGSPTEVGGEIEKLRKERR
ALMEEMVELQQQSRGTARHVDVTNQRLLKAAEQRQKQLLSFLAKLFQNRGFLERLKNFKGK
EKGALGLEKARKKFIKHQQPQDSPTGGEVVKYEADDWERLLMYDEETENTKGLGGMTS
SDPKGKNLMYPSEEEEMSKPDYLSFSPPEGLIKQEETWMSMGFDTTIPSFNTDAWGNTM
DYNDVSEFGFAAETTS DGLPDVCWEQFAAGITETGFNWPTGDDDDNTPMNDP*
>G2094 (1..450)
ATGCTAGATCCCACGAGAAAGTAATCGATTGAGAAATCAATGGAAAGCAAACCTCACATCA
GTAGATGCGATCGAAGAACACAGCAGCAGTAGCAGTAATGAAGCTATCAGCAACGAGAAG
AAGAGTTGTGCCATTTGTGTGTTACAGCAAAACCCCTCTTGGCGAGGCGGTCTGCCGGT
CCCAAGTCGCTTTGTAACGCATGCGGGATCAGAAACAGAAAGAAAGAAAGAACTGATC
TCAAATAGATCAGAAGATAAGAAGAAGAGTCAATACAGAAACCCGAAGTTTGGTGAC
TCGTTGAAGCAGCGATTAAATGGAATTGGGGAGAGAAGTGATGATGCAGCGATCAACGGCT
GAGAATCAACGGCGGAATAAGCTTGGCGAAGAGCAAGCCGCGTGTACTCATGGCT
CTCTCTTATGCTTCTCCGTTTATGCTTAA

>G2094 Amino Acid Sequence (domain in AA coordinates:43-68)
MLDPTEKVIDSESMESKLTSDVAIEEHSSSSSNEAISNEKKSCAICGTSKTPLRGGGPAG
PKSLCNACGIRNRKRRTLISNRSEDKKKKSHNRNPKFGDSLKQRLMELGREVMMQRSTA
ENQRRNKLGEEEQAAVLLMALSYASSVYA*

>G2113 (90..590)
ATAACAAACTCATCAAACCTCCTCAGCGTTTCTTTTCTTACATAAACAATTTTCTTAC
ATAAACAAATCTTGTGTTTGTGTGTGTCATGGCACCAGACAGTTAAACCGCGGCCGTCA
AAACCAACGAAGGTAACGGAGTCCGTTACAGAGGAGTGAGGAAGAGACCATGGGGACGTT
ACGCAGCCGAGATCAGAGATCCTTTCAAGAAGTCACGTGTCTGGCTCGGTACTTTTCGACA
CTCCTGAAGAAGCCGCTCGTGCCTACGACAAACGTGCTATTGAGTTTTCGTGGAGCTAAAG
CCAAAACCAACTTCCCTTGTGTACAACATCAACGCCCACTGCTTGAGTTTGACACAGAGCC
TGAGCCAGAGCAGCACCGTGGAATCATCGTTTCTAATCTCAACCTCGGATCTGACTCTG
TTAGTTCGAGATTCCCTTTTCTTAAGATTGAGTTAAGGCTGGGATGATGGTGTTCGATG
AAAGGAGTGAATCGGATTCTTCGTGCGTGGTGTGATGGATGTCGTTAGATATGAAGGACGAC
GTGTGGTTTGGACTTGGATCTTAATTTCCCTCCTCCACCTGAGAACTGATTAAGATTTA
ATTATGATTATTAGATATAAATTAAATGTTTCTGAATTGAG

>G2113 Amino Acid Sequence (domain in AA coordinates: TBD)
MAPTVKTAAVKTNEGNGVRYRGVGRKPWGRYAAEIRDPFKSRVWLGTFTDPEEAAARYD
KRAIEFRGAKAKTNFPCYNINAHCLSLTQSLQSSTVSESSFPNLNLGSDSVSSRFPFPI
QVKAGMMVFDERSESDSSSVMDVVRYEGRRVVLDDLNLNFPFPPEN*

>G2115 (41..733)
AATCACTCTACAAAGCCTGTACGTACACAACAACATTACCATGGTGAAACAAGAACGCAA
GATCCAAACCAGCAGCACAAAAAGGAAATGCCTTTGTTCATCATCACCATCTTCTTCTTC
TTCTTTCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
TAAGAAGTACAAAGGAGTGAGGATGAGAAGTTGGGGATCATGGGTCTCTGAGATTAGGGC
ACCAATCAAAAGACAAGGATTGTTAGGTTCTTACTCAACAGCTGAAGCAGCTGCTAG
AGCTTACGATGTTGCACTCTTATGTCTCAAAGGCCCTCAAGCCAATCTCAACTTCCCTAC
TTCTTCTTCTTCTCATCATCTTCTTGTATAATCTCTTAGATGAAAATACCCCTTTGTCCCC
CAAATCCATCCAAAGAGTAGCTGCTCAAGCTGCCAATCATTTAACCATTTTGCCCTAC
TTCATCAGCCGCTCTCGTCACCGTCCGATCATGATCATCACCATGATGATGGGATGCAATC
TTTGATGGGATCTTTTGTGGACAATCATGTGTCTTTGATGGATTCAACATCTTCATGGTA
TGATGATCATAATGGGATGTTCTTGTGTGATAATGGAGCTCCATTCAATTACTCTCCTCA
ACTAAACTCGACGACGATGCTCGATGAATACTTCTACGAAGATGCTGACATTCCGCTTTG
GAGTTTCAATTAATCCGACGGTCCATAATACATACTTTAATTAGT

>G2115 Amino Acid Sequence (conserved domain in AA coordinates:46-115)
MVKQERKIQTSSTKKEMPLSSSSSSSSSSSSSSSSSSCKNKNKSKIKKYKVRMRSWGS
WVSEIRAPNQKTRIWLGSYSTAEAAARAYDVALLCLKGPQANLNFTSSSSHLLDNLDD
ENTLLSPKSIQRVAAQAANSFNHFAPTSSAVSSPSDHDHHDHDDGMQSLMGSFVDNHVSLM
DSTSSWYDDHNGMFLFDNGAPFNYSPLNSTTMLDEYFYEDADIPLWSFN*

>G2130 (41..988)
CCTCTCTTCATTTTAACTCCCTCTCTCTCTCTCTCTATGGAGAGACGAACGAGACG
AGTGAAGTTCACAGAGAATCGTACGGTCACAAACGTAGCAGCTACACCATCTAACGGGTC
TCCGAGACTGGTCCGTATCACTGTTACTGATCCTTTTCGCTACTGACTCGTCTAGCGACGA
CGACGACAACAACCGTCACGGTGGTTCCAAGAGTGAAACGATACGTGAAGGAGATTAG
ATTCTGCCAAGGTGAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
GGAAAGCGTAGTGGTTGAAGATGACGTGTGACGTGCGGTGAAGCCTAAAAAGTACAGAGG
CGTGAGACAGAGACCTTGGGGAAAATTGCGCGCGGAGATTAGAGATCCGTCGAGCCGTAC
TCGGATTTGGCTTGGGACTTTTGTACGGCGGAGGAAGCTGCTATAGCGTACGATAGAGC
GCCGATTCTCTCAAAGGACCTAAAGCGCTCACGAATTTCTTAATCTCCGCCGACGCCAAC
CCCGGTTATCGATCTCCAAACGGTTTCCGCTGCGATTACGGTAGAGATTCTCGGCAGAG
CCTTCATTACCCGACCTCTGTTCTAAGATTCAACGTCAACGAGGAAAACAGAGCATGAGAT
TGAAGCGATCGAGCTATCTCCGAGAGAAAGTGCACGGTTATAAAAGAAGAAGAAGATC
GTCGGCGGGTTTGGTGTTCCTGGATCCGTATCTGTTACCGGATTTATCTCTCGCCGCGCA
ATGTTTTTGGGATACCGAAATTGCCCTGACCTTTTGTCTCGATGAAGAAACCAAAAT
CCAATCAACGTTGTTACCAAAACACAGAGGTTTCGAAACAAGGAGAAAACGAAACTGAAGA
TTTCGAGTTTGGTTGATTGATGATTTTCGAGTCTTCTCCATGGGATGTGGATCATTTCTT
CGACCATCATCACTCTTTCGATTAAAAATCTCTTCTTTTGGGGAAATTTTGTG

>G2130 Amino Acid Sequence (domain in AA coordinates 93-160)
MERRTRRVKFTENRTVTNVAATPSNGSPRLVLRITVTDPFATDSSSSDDDDNNNVTVVPRVK
RYVKEIRFCQGESSSSSTAARKGKHKEESVVVEDDVSTSVKPKKYRGVRQRPWGKFAAEI
RDPSSRTRIWLGTFTVTAEEAAIAYDRAAIHLKGPALTNFLTPTPTPTVIDLQTVSACDY
GRDSRQSLHSPTSVLRFNVNEETEHEIEAIELSPERKSTVIKEEEESSAGLVFPDPYLLP
DLSLAGECFWDTEIAPDLLFLDEETKIQSTLLPNTEVSKQGENETEDFEFGLIDDFESSP
WDVDHFFDHHHHSFD*

>G2147 (162..1262)

CTGTGATTGTCAAGAGTTTGAACACACAAAGAAGAAAGAAGAACTCAACATTTCAAGCAA
GAAGAAAGAGAGAAGAGAGAAGGTCCAATAATAGAGAGAACAAAAAAGAGAGCTTAA
TTGTTCAGTTTATTCTCTGCAACCGTGCAGGCTTAAGTAACACATGTCTGAATTTATGGAGTTA
AAGAGCTCACATGGGAAAATGGGCAACTAACCGTTTCATGGTCTAGGCGACGAAGTAGAAC
CAACCACCTCGAATAACCCCTATTGGACTCAAAGTCTCAACGGTTGTGAGACTTTGGAGT
CTGTGGTTTCATCAAGCGGCTCTACAGCAGCCAAGCAAGTTTCAGCTGCAGAGTCCGAATG
GTCCAAACCACAATTATGAGAGCAAGGATGGATCTTGTTCAGAAAACCGCGTTATCCTC
AAGAAATGGACCGATGGTTCGCTGTTCAGAGGAGAGCCATAGAGTTGGCCACAGCGTCA
CTGCAAGTCCGAGTGGTACCAATATGTCTTGGGCGTCTTTTGAATCCGGTCCGAGCTTGA
AGACAGCTAGAACCGGAGACAGAGACTATTTCCGCTCTGGATCGGAACTCAAGATACTG
AAGGAGATGAACAAGAGACAAGAGGAGAAGCAGGTAGATCTAATGGACGACGGGGACGAG
CAGCAGCGATTACAACGAGTCCGAAAGGAGACGGCGTGATAGGATAAACCAGAGGATGA
GAACACTTCAGAAGCTGCTTCTACTGCAAGTAAGGCGGATAAAGTCTCAATCTTGGATG
ATGTTATCGAACACTTGAACAGCTACAAGCACAAGTACAGTTTCATGAGCCTAAGAGCCA
ACTTGCCACAACAATGATGATTCCGCAACTACCTCCACCACAGTCAGTTCTCAGCATCC
AACACCAACAACAACAACAACAGCAGCAGCAGCAGCAACAACAGCAGCAACAGTTTC
AGATGTCGTTGCTTGCAACAATGGCAAGAATGGGAATGGGAGGTGGTGGAAATGGTTATG
GAGGTTTAGTTCTCTCTCTCTCTCTCCACCAATGATGGTCCCTCTCTATGGGTAACAGAG
ACTGCACCAACGGTCTCTTCAGCCACATTATCTGATCCATACAGCGCTTTTTCGCACAGA
CAATGAATATGGATCTCTACAATAAATGGCAGCAGCTATCTATAGACAACAGTCTGATC
AAACAACAAGGTAAATATCGGCATGCCTTCAAGTTCTTCAAGTTCATGAGAAAAGAGATT
AGTCTAGCGACCTAGTATTATTGATCCATATATATAGTTCTTGAAAGATTGTTGTATCAT
GATTGTA AAAACTGTTTTGAGTATGGA AAAAGACTTG CAGATAAAA

>G2147 Amino Acid Sequence (domain in AA coordinates:160-234)
MSNYGVKELTWENGQLTVHGLGDEVEPTTSNNPIWTQSLNGCETLESVVHQAALQQPSKF
QLQSPNGPNHNYESKDGSCSRKRGYPOEMDRWFAVQEEHVRVGHSVTASASGTNMSWASF
ESGRSLKTARTGDRDYFRSGSETQDTEGDEQEETRGEAGRSNGRRGAAAAIHNESERRRRD
RINQRMRTLQKLLPTASKADKVSILDDVIEHLKQLQAQVQFMSLRANLPQQMMIPQLPPP
QSVLSIQHQQQQQQQQQQQQQQQQFQMSLLATMARMGMGGGNGYGGGLVPPPPPPPMV
PPMGNRDCTNGSSATLSDPYSAFFAQTMNMDLYNKMAAAIYRQSDQTTKVNIGMPSSSS
NHEKRD*

>G2156 (384..1292)

TTTTTTTTCCCTTTCTCGTTCAAAAAAGTACTTGCAGAGTCACTCACTCTCAGTCTCA
GCACATGAATTAATTTGAAGCTTCCCTAGAATTCTTTACATCAATTAATACGACACCGT
CTCGGGTGAAGAATCTCTCTCTCTTGGCCCTAAAGCGAGTTAGGGTTTAACACACAAAGC
ATACCCTTTAGATTGTGTCTCTTAGCTCTGTTTTGTGCGGCTTGTGTAACCGATCAACT
CAAGCTATTGGCTCTCACCTCCTGAAATTTGACTTCTCCAATGGATCTCAAAGTTTCTC
TTATATGAATTCTATCTTCACTTCAATATCTTTATATATATGAGCCACAAGAACAAG
AAGAGTCAGTAGATGCGGCTGCCATGGACGGTGGTTACGATCAATCCGGAGGAGCTTCTA
GATACTTTCACAACCTCTTCAGGCTGAGCTTCATCACCAGCTTCAACCTCAGCCTCAAC
TTCACCTTTGCTCAGCCTCAGCCTCAACCTCAGCCTCAGCAGCAGAATTCAGATGATG
AATCTGACTCCAACAAGGATCCGGGTTCCGACCCAGTTACCTCTGGTTCAACCGGGAAC
GTCCACGTGGACGTCTCCGGGATCCAAGAACAAGCCGAAGCCACCGGTGATAGTACTA
GAGATAGCCCCAACGTGCTTAGATCTCATGTTCTTGAAGTCTCATCTGGAGCCGACATAG
TCGAGAGCGTTACCATCTACGCTCGCAGGAGAGGAAGAGGAGTCTCCATTCTCAGTGGTA
ACGGCACGGTGGCTAACGTGAGTCTCCGGCAGCCGGCAACGACAGCGGCTCATGGGGCAA
ATGGTGAACCGGAGGTGTTGTGGCTCTACATGGAAGGTTTGAGATACTTTCCCTCACAG
GTACGGTGTGTCGCCCCCTGCGCCGCCAGGATCCGGTGGTCTTTCTATCTTTCTTCCG
CCGTTCAAGGTCAGGTGATTGGAGGAAACGTGGTGGCTCCGCTTGTGGCTTCGGGTCCAG

AACGGTGAAGGCCGATGGAATCTGCTCGCTAAATCTTCTGGGCTAAAGAGAGCAGGAAAA
AGTTGTAGATTGAGATGGTTGAATTACCTTAAACCCGACATAAAGCGTGGAATCTCACT
CCTCAAGAACAACCTTTAATCCTTGAGCTCCATTCTAAATGGGGTAATAGGTGGTCAAAA
ATTTCAAGTATTTACCAGGAAGAACAGACAACGATATCAAAACTACTGGAGAAGTAGA
GTCCAGAAACAAGCAGCCAGCTCAACATAGATTCCAATAGCCACAAGTTCATAGAAGTT
GTTCTAGCTTTTGGTTTCCAAGACTGATCAACGAGATTAAAGACAACCTCATACACCAAC
AATATTAAGCTAATGCTCCTGATTTACTTGGACCAATTTTACGAGACAGCAAAGATTG
GGTTTCAACAACATGGATTGTTCCACTTCCATGTCAGAAGATCTCAAGAAAACCTCACAA
TTCATGGATTTTCTGATCTTGAACCACAATGTCCTTGAAGGATCACGAGGGGGTAGT
AGTCAATGTGTGAGTGAGGTTTATAGCTCCTTCCCTTGCCTAGAGGAGGAGTACATGGTG
GCCGTTATGGGCAGTTCAGACATTTTCAGCATTCATGATTGTCACGTGGCTGATTCCAAG
TACGAGGATGATGTGACACAAGATCTAATGTGAACATGGATGACATTTGGCAGTTTAAAC
GAGTATGCACACTTTAATTAGGTTATATTATTTATGTACTTCTTACAACCTGGAGGGG
TTTATCGGTCTTTTATTAAATTTTGATTGTTTGGATTCTTAAAAATGTGTTCTTATTA
TAGTTTTTAATGAAAAAATGTTTAAAGCGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
>G2893 Amino Acid Sequence (conserved domain in AA coordinates:19-120)
MSNITKKKCNNGNEEGAEQRKGPWTLEEDTLLTNYISHNTEGRWNLAKSSGLKRAGKSCR
LRWLNLPDIKRGNLTPQEQLLILELHSGWNRWSKISKYLPGRDNDIKNYWRTRVQK
QARQLNIDSNSHKFIEVVRSFWFPRLINEIKDINSYNNIKANAPDLLGPILRDSKDLGFN
NMDCSTSMSEDLKKTQSFMDFSLETMSLESGRGSQCVSEVYSSFPCLSEYVMVAVM
GSSDISALHDCHVADSKYEDDVTQDLMWNMDDIWQFNEYAHFN*
>G340 (97..834)
ATGAAATCTCTGTAGTTTTTTTTTGTTCCTTTCTTAAATTTTCAAAGAAAGACATTTATT
AAACCAAAATAACTCTTTAGATCATTGCAAGGAAAAATGTTGAAAAGTGCAAGTCCAATG
GCATTCTACGATATCGGAGAGCAGCAATACTCTACTTTTCGGGTACATTTTAAGCAAACCT
GGGAACGCAGGAGCTTACGAGATTGACCTTCGATCCCAAACATCGACGATGCGATCTAC
GGCTCAGATGAGTTCCGTATGTACGCTTACAAAATCAAACGGTGTCTCTGACTCTGATGC
CACGACTGGACGGAGTGTCCTTACGCTCACCCTGGCGAGAAAGCCACACGCCGTGATCCT
CGCCGTTACACTTACTGTGCGATGCGATGCCCCGCTTTCCGAAATGGCGCATGCCACCGT
GGCGACTCATGCGAATTCGCACATGGCGTATTCGAGTACTGGCTCCACCCGGCGCGTTAC
CGAACACGCGCATGTAACGCCCGGGAACCTTGTGTGAGGAAAGTGTTTCTTTGCCAC
GCGCCGGAGCAGCTAAGGCAGTCTGAAGGAAAGCACAGGTGCAGGTACGCATATAGGCCG
GTGAGGGCTAGAGGTGGTGAAACGGCGATGGAGTGACGATGAGAATGGACGACGAGGGT
TAGCACACGTACGGTCTCCGGTGAGAAGCGGGAAGATGATTTAGATAGTAACGAGGAG
AAGTGTTTGTGAAGTGTGGAGTCCGATGAGCATTGTGGATGATCATTATGAGCCGTCC
GATTTGGATTTGGATTTGTCACTTTGATTGGATCTCAGAGTTGGTCGATTAAATTTGG
GAAATCAAAGCAGAGAAACAAAAGAAACCCGATAAATAAAGTGGATTTTGTGTTAAATCCAC
AAGATCAAGATTCAAGATGAGAGATCTTGTGATGATATGGTAAATTTAATTGTAATGAT
TTATTGCAATGTCGCAAAAGAAGTTACTTCTCTTTCATGTAAACAGATTCTTGATCTTC
TATAAGTCTTTGTATTAA
>G340 Amino Acid Sequence (domain in AA coordinates: 37-154)
MLKSASPMFYDIGEQYSTFGYILSKPGNAGAYEIDPSIPNIDDAIYGSDEFMYAYKI
KRCPRTRSHDWTECPYAHRGEKATRRDPRRYTYCAVACPFRNGACHRGDSCEFAHGVFE
YWLHPARYRTRACNAGNLCQRKVCFFAHAPEQLRQSEGHRCRYAYRPVRARGGGNGDGV
TMRMDDEGYDTSRSPVRSGKDDLSNEEKVLLKCWSRMSIVDDHYEPSDLDDLDSHFDWI
SELVD*
>G39 (75..638)
GTTTCCACAGTCCCTGTACTTGTGCATAAACTGTAAACACTACTCTGAAAATTTTGCT
TCTGTTAGGATATAATGCCACCCTCTCTCTAAATCTCCTTTTATTAGCTCTTCACTCA
AAGGAGCTCATGAAGATCGCAAATTTAAATGCTATAGGGGTGTCCGAAAGAGGTCTTGGG
GCAAATGGGTGTCTGAAATCAGAGTTCCAAAGACTGGACGACGAATATGGCTAGGTTTAT
ACGATGCTCCAGAGAAGGCAGCTAGAGCCTATGATGCTGCTTTGTTCTGTATTAGGGGTG
AGAAGGGAGTTTACAATTTTCCCACTGATAAAAAGCCGAGCTTCAGAAAGGTTCTGTCC
GGCCTCTGTCCAAGCTCGACATACAGACAATAGCAACAACTATGCTTCATCAGTTGTGC
ATGTACCTTCCCATGCCACCACACTCCCGGCAACAACCCAGGTTCCCTCTGAAGTTCCTG
CTTCCTCTGATGTTTCTGCTTCTACTGAGATTACAGAGATGGTCGATGAATATTATCTCC
CAACCGATGCAACTGCAGAATCAATATTCTCAGTTGAAGACTTACAACCTGGACAGTTTCC

TCATGATGGACATTGATTGGATAAACAATCTAATCTGATGTGTAACGTCACCTGCAGTGA
CATTTAATATGGTTTANCTATCAGTTACCTGTCTGCTTCTTGTAAGGGTATACTTGGATC
CTTGCTCTTGAACCTGTTTTATTAGCATGCAAA

>G39 Amino Acid Sequence (domain in AA coordinates: 24-90)

MPPSPKSPFISSSLKGAHEDRKFKCYRGVRKRSWGKVVSEIRVPKTGRRIWLGSYDAPE
KAARAYDAALFCIRGEKGVNFPDCKPQLPEGSVRPLSKLDIQTATNYASSVVHVP
SH
ATTLPATTPVPSEVPASSDVSASTEITEMVDEYLPD
DATAESIFSVEDLQLDSFLMMDI
DWINNLI*

>G439 (128..967)

TATAAATCTTCGTTTCTACTTTTTTTTCTTCCATAATATAGTCAATTCGTTTCTTAATT
AGGGCTTCTTCTCTTTGTTTCTCCAATCTTTATTAGTTTATTTATTTATTTTGGTTATTG
TATACAAATGGCAATGGCTTTAAACATGAATGCTTACGTAGACGAGTTCATGGAAGCTCT
TGAACCATTCATGAAGGTAACCTCATCTTCTTCTACTTCAATTCATCAAAATCCAAAACC
ATTAATCCTTAATTTTATCCCTAATAATGACCAAGTCTTACCGGTATCTAACCAAACCGG
TCCGATTGGGCTAAACCAGCTCACTCCAACACAAATCCTCCAAATTCAGACAGAGTTACA
TCTCCGGCAAAACCAATCTCGTCGTCGCGCTGGTAGTCATCTTCTCACCCTAAACCAAC
CTCAATGAAGAAAATCGACGTAGCAACTAAACCGGTTAAACTATACCGAGGCGTAAGACA
GAGGCAATGGGGTAAATGGGTAGCTGAGATTGCGCTACCTAAAAACCGAACCCGGTTATG
GCTCGGTACGTTGCAAAACGGCTCAAGAAGCTGCATTAGCTTACGATCAAGCAGCTCATAA
GATCAGAGGAGACAACGCTCGTCTCAATTTCCAGACATTGTTGCTCAAGGACACTATAA
ACAGATATTGTTCTCGTCTATCAACGCAAGATCGAATCCATCTGCAATAGTTCTGATCT
TCCACTGCCTCAGATCGAGAAACAGAACAAAACAGAGGAGGTGCTCTCTGGTTTTCCTAA
ACCGGAGAAAGAACCGGAATTTGGGGAGATATACGGATGCGGATACTCGGGCTCATCTCC
TGAGTCGGATATAACGTTGTTGGATTTCTCAAGCGACTGTGTGAAAGAAGATGAGAGTTT
CTTGATGGGTTTGACAAGTATCCTTCTTTGGAGATTGATTGGGACGCTATAGAGAACT
CTTCTGAATCCATTTTATCTTTTTTGATTTCATTTGTCTCTAAATTGTAGAATTTTATTTTC
AGAGCTTTGTAAGGGAAGTTCTTGAATGAGAGTTGCAGAGGACTAGTGGAACCTAACTCT
GTTTTCTTTTGTAAATGATTGTTTATAATGGGCCGTTGAATGGGCCTTATTGATTTAAACA
GCCCAAGTTTTTAAAAA
AAAAAAAAAAAAAAAAAAAAA

>G439 Amino Acid Sequence (domain in AA coordinates: 110-177)

MAMALNMNAYVDEFMEALEPFMKVTSSTSSSNPKPLTPNFI PNNDQVLPVSNQTGPI
GLNQLTPTQILQIQTELHLRQNSRRRAGSHLLTAKPTSMKKIDVATKPVKLYRGVRQRQ
WGKWVAEIRLPKNRRLWLGTFFETAQEAALAYDQAHKIRGDNARLNFPDIVRQGHYKQI
LSPSINAKIESICNSSDLPLPQIEKQNKTEBVLGSGFSKPEKEPEFGEIYCGYSGSSPES
DITLLDFSSDCVKEDESFLMGLHKYPSLEIDWDAIEKLF*

>G470 (1..2580)

ATGGCGAGTTTCGGAGGTTTCAATGAAAGGTAATCGTGGAGGAGATAACTTCTCCTCCTCT
GGTTTTAGTGACCCTAAGGAGACTAGAAATGTCTCCGTCGCCGCGGAGGGGCAAAAAGT
AATTTCTACCCGATCCGCTGCGGCTGAGCGTGCTTTGGACCTGAGGCTGCTCTTTACAGA
GAGCTATGGCACGCTTGCTGCTGGTCCGCTTGTCACGGTTCCTAGACAAGACGACCGAGTC
TTCTATTTTCTCAAGGACACATCGAGCAGGTGGAGGCTTCGACGAACAGGCGGCAGAA
CAACAGATGCCTCTCTATGATCTTCCGTCAAAGCTTCTGTCGAGTTATTAATGTAGAT
TTAAAGGCAGAGGCAGATACAGATGAAGTTTATGCGCAGATTACTTCTTCTCCTGAGGCT
AATCAAGACGAGAATGCAATGAGAAAGAAGCGCCTCTTCTCCACCTCCGAGGTTCCAG
GTGCATTGCTTCTGCAAAACCTTGACTGCATCCGACACAAGTACACATGGTGGATTTTCT
GTTCTTAGGCGACATGCGGATGAATGTCTCCACCTCTGGATATGTCTCGACAGCCTCCC
ACTCAAGAGTTAGTTGCAAAAGATTGTCATGCAAAATGAGTGGCGATTTCAGACATATATTC
CGGGGTCAACCACGSAAGGCAATTTGCTACAGAGTGGGTGGAGTGTGTTGTTAGCTCCAAA
AGGCTAGTTGCAGGCGATGCGTTTATATTTCTAAGGGGCGAGAATGGAGAATTAAGAGTT
GGTGTAAAGGCGTGGATGCGACAACAAGGAAACGTGCCGTCTTCTGTTATATCTAGCCAT
AGCATGCATCTTGAGTACTGGCCACCGCATGGCATGCCATTTCACAGGGACTATGTTT
ACAGTCTACTACAAACCCAGGACGAGCCCATCTGAGTTTATTGTTCCGTTTCGATCAGTAT
ATGGAGTCTGTTAAGAATAACTACTCTATTGGCATGAGATTCAAATGAGATTTGAAGGC
GAAGAGGCTCCTGAGCAGAGGTTTACTGGCACAATCGTTGGGATTGAAGAGTCTGATCCT
ACTAGGTGGCCAAATCAAAGTGGAGATCCCTCAAGGTGAGATGGGATGAGACTTCTAGT
ATTCTCGACTGATAGAGTATCTCCGTGGAAAGTAGAGCCAGCTCTTGCTCCTCCTGCT
TTGAGTCCTGTTCCAATGCCTAGGCCTAAGAGGCCAGATCAAATATAGCACCTTCATCT

CCTGACTCTTCGATGCTTACCAGAGAAGGTACAACCTAAGGCAAACATGGACCCCTTTACCA
GCAAGCGGACTTTCAAGGGTCTTGCAAGGTCAAGAATACTCGACCTTGAGGACGAAACAT
ACTGAGAGTGTAGAGTGTGATGCTCCTGAGAATTCTGTTGTCTGGCAATCTTCAGCGGAT
GATGATAAGGTTGACGTGGTTTCGGGTTCTAGAAGATATGGATCTGAGAACTGGATGTCC
TCAGCCAGGCATGAACCTACTTACACAGATTTGCTCTCCGGCTTTGGGACTAACATAGAT
CCATCCCATTGGTCAGCGGATACCTTTTTATGACCATTCATCATCACCTTCTATGCCTGCA
AAGAGAATCTTGAGTGATTGAGAAGGCAAGTTCGATTATCTTGCTAACCACTGGCAGATG
ATACACTCTGGTCTCTCCCTGAAGTTACATGAATCTCCTAAGGTACCTGCAGCAACTGAT
GCGTCTCTCCAAGGGCGATGCAATGTTAAATACAGCGAATATCCTGTTCTTAATGGTCTA
TCGACTGAGAATGCTGGTGGTAACTGGCCAATACGTCCACGTGCTTTGAATTATTATGAG
GAAGTGGTCAATGCTCAAGCGCAAGCTCAGGCTAGGGAGCAAGTAACAAAACAACCCCTC
ACGATACAAGAGGAGACAGCAAAGTCAAGAGAAGGGAAGTGCAGGCTCTTTGGCATTCTCT
CTGACCAACAACATGAATGGGACAGACTCAACCATGTCTCAGAGAAACAACCTGAATGAT
GCTGCGGGGCTTACACAGATAGCATCACCAAAGGTTCAAGACCTTTCAGATCAGTCAAAA
GGGTCAAAATCAACAAACGATCATCGTGAACAGGGAAGACCATTCCAGACTAATAATCCT
CATCCGAAGGATGCTCAACAGAAAACCAACTCAAGTAGGAGTTGCACAAAGGTTCAACAG
CAGGGAATTGCACTTGGCCGTTCACTGGATCTTTCAAAGTTCAAAACCTATGAGGAGTTA
GTCGCTGAGCTGGACAGGCTGTTTGAAGTCAATGGAGAGTTGATGGCTCCTAAGAAAGAT
TGGTTGATAGTTTACACAGATGAAGAGAATGATATGATGCTTGTGTTGGTGACGATCCTTGG
CAGGAGTTTGTGTCATGGTTCGCAAAATCTTCATATACAGAAAGGGAAGTGAGGAAG
ATGAACCCGGGGACTTTAAGCTGTAGGAGCGAGGAAGAAGCAGTTGTTGGGGAAGGATCA
GATGCAAGGACGCCAAGTCTGCATCAAATCCTTCATTGTCCAGCGCTGGGAACCTCTTAA
>G470 Amino Acid Sequence (domain in AA coordinates: 61-393)
MASSEVSMKGNRGGDNFSSSGFSDPKETRNVS VAGEGQKSNSTRSAAERALDPEALYR
ELWHACAGPLVTVPRQDDRVFYFPQGHIEQVEASTNQAAEQMPLYDLPSKLLCRVINVD
LKAEDTDEVYAQITLLPEANQDENAIKEAPLPPPPRFQVHSFCKLTASDTSTHGGFS
VLRRADECLPPLDMSRQPPTQELVAKDLHANEWRFRIFRGQPRRHLLQSGWSVVFSSK
RLVAGDAFIFLGRGELRVGVRRAMRQQGNVPSSVISSSHMLGLVATAWAHAISTGTMF
TVYYKPRTPSEFIVPFDQYMESVKNNYSIGMRFKMRFEGEAEPEQRFTGTIVGIEESDP
TRWPKSKWRSLKVRWDETSSIPRDRVSPWKVEPALAPPALSPVPMRPRKRPRSNIAPSS
PDSSMLTREGTTKANMDPLPASGLSRVLQGGQYESTLRTKHTESVECDAPENSVVWQSSAD
DDKVDVVS GSRRYGSENWMS SARHEPTYD LLSGFGTNIDPSHGQRIPFYDHSSSPMPA
KRILSDSEGKFDYLANQWQMIHSGLSLKLHESPKVPAATDASLQGRCNVKYSEYPVLNGL
STENAGGNWPIRPRALNYEEVNAQAQAQAREQVTKQPTTQIETAKSREGNCRFLGIP
LTNNMNGTDSTMSQRNNLNDAGLTQIASPKVQDLSDQSKGSKSTNDHREQRPFQTNPN
HPKDAQTKTNSRSTCKVHKQGIALLGRSVDLSKFQNYEELVAELDRLFEFNGELMAPKKD
WLI VYTDEENDMLVGD DPWQEFCCMVRKIFIYTKEEVRKMNPGT LSCRSEEEAVVGE
DAKDAKSASNPSLSAGNS*

>G652 (1..606)

atgagcggaggagagacgtgaacatgagtggtggagacagacgcaagggaaacggtgaag
tggtttgatacacagaaggggttggtttcatcacacctagcgacggtggtgacgatctc
tcggttcaccagctctccatcagatctgaaggatttcgtagcctcgcagctgaggaatct
ggtgagttcgacgttgaggttgacaactccggccgtcccaaggctattgaagtgtctgga
cccagcgggtgctcccgttcagggttaacagcgggtggtggtggttcatctggtggacgcgg
gggttttgccggcggtggtggaagaggagggggacgtggtggaggaagctacggaggaggt
tatggtggaagaggaagcgggtggcgtggaggaggtggtggtgataattcttgctttaag
tgcggtgaaccaggtcacatggcgagagaatgctctcaagggtggtggaggatacagcgg
ggcggggggtggtggaaggtacgggtctggcggcgcgaggaggaggtggtggtggctta
agctgctacagctgtggagagctctgggcactttgcaagggttgactagcgggtggtgct
cgttga

>G652 Amino Acid Sequence (domain in AA coordinates: 28-49, 137-151, 182-196)

MSGGGDVNMSGGDRRKGTVKWFDTQKGFITPSDGGDDL FVHQSSIRSEGFRSLAAEES
VEFDVEVDNSGRPKAIEVSGPDGAPVQGNSSGGGSSGGRGFGGGGGRGGRGGGSGYGGG
YGGRGSGRGGGGGDNSCFKCGEPGHMARECSQGGGGYSGGGGGGRYSGGGGGGGGGGL
SCYSCGESGHFARDCTSGGAR*

>G671 (61..1119)

TTCACTTGAGAAACACCCCTTTGAACTCGATCAAGAAAGCTAAGTTTGAAGAATCAAGA

ATGGTGCGGACACCGTGTGCAAAGCCGAAGTGGGTAAAGAAAGGAGCTTGGACTCCC
GAGGAAGATCAGAAGCTTCTCTTACCTTAACCGCCACGGTGAAGGTGGATGGCGAAGT
CTCCCCGAAAAAGCTGGACTCAAGAGATGCGGCAAAAGCTGCAGACTGAGATGGGCCAAT
TATCTTAGACCTGACATCAAAAGAGGAGAGTTCAGTGAAGACGAAGAAGCTTCAATCATC
TCTCTTACGCCCCCTTACCGGCAACAAATGGTCTGCTATAGCTCGTGGACTACCAGGAAGA
ACCGATAACGAGATCAAGAAGTACTGGAACACTCATATCAAAAAACGTTTGATCAAGAAA
GGTATTGATCCAGTTACACACAAGGGCATAACCTCCGGTACCGACAAATCAGAAAACTC
CCGGAGAAACAAAATGTTAATCTGACAACTAGTGACCATGATCTTGATAATGACAAGGCG
AAGAAGAACAACAAGAATTTTGGATTATCATCGGCTAGTTTCTTGAACAAAGTAGCTAAT
AGGTTTCGGAAGAGAATCAATCAGAGTGTCTGTCTGAGATTATCGGAAGTGGAGGCCCA
CTTGCTTCTACTAGTCACTACTAATACTACAACCTACAAGTGTTCCTGTTGACTCTGAA
TCAGTTAAGTCAACGAGTTCCTTCCCTTCGCACCAACCTCGAATCTTCTCTGCCATGGGACC
GTTGCAACAACACCAAGTTTCATCGAAGTTTGACGTTGATGGTAACGTTAATCTGACGTGT
TCTTCGTCACGTTCTCTGATTCTCCGTTAACAATCCTCTAATGTACTGCGATAATTTTC
GTTGGTAATAACAACGTTGATGATGAGGATACTATCGGGTCTCCACATTTCTGAATGAT
GAAGATTTTCATGATGTTGGAGGAGTCTTGTGTTGAAAACACTGCGTTCATGAAAGAACTT
ACGAGGTTTCTTCACGAGGATGAAAACGACGTCGTTGATGTGACGCCGGTCTATGAACGT
CAAGACTTGTTTGACGAAATTGATACTATTTTGGATGAGTGAAACTCATAATCGATGAA
TCCCACGTGACCATGTCAATATGATGTCTATGGATATGTTACCTTGATGATGTTGATGGT
AATAATAATAAATAATAGATGGTGTATGATGACCATGCATGAATCATGAATGTAGTTCGTG
TTGTCACATATGCTTGTGTTTTTGTGTTTTTTTTTTTGGTCTGAAGTGTGTTGTTTCGT
TGTAATGATTATAAATGGTGTATGTAATAATTATAATGTTAAAAAATAAATAAATAA
AAAA

>G671 Amino Acid Sequence (domain in AA coordinates: 15-115)
MVRTPCCKAELGLKKGAWTPEEDQKLLSYLNRHGEQWRTLPKAGLKRCGKSCRLRWAN
YLRPDIKRGEFTEDEERSIISLHALHGNKWSAIARGLPGRDNEIKNYWNTHIKRLIKK
GIDPVTHKGITSGTDKSENLEKQNVNLTSDHDLNDKAKKNNKNFGLSSASFLNKVAN
RFGKRINQSVLSEIIGSGGGLASTSHTTNTTTSVSVDSSESVKSTSSSFAPTSNLLCHGT
VATTPVSSNFDVDGNVNLTCSSSTFSDSSVNPLMYCDNFVGNVNDDEDITIGFSTFLND
EDFMMLEESCVENTAFMKELTRFLHEDENDVDVTPVYERQDLFDEIDNYFG*

>G779 (110..712)

GACATGCATGTAAGCAATTCGGTTAATTAATCGAGTCAAAGATATATATCAGTAAATACAT
ATGTGTATATTTCTGAAAAAGAATATATATATTGAGAAATAAGAAAAGATGAAAATGGA
AAATGGTATGTATAAAAAGAAAGGAGTGTGCGACTCTTGTGTCTCGTCCAAAAGCAGATC
CAACCACAGCCCCAAAAGAAGCATGATGGAGCCTCAGCCTCACCATCTCCTCATGGATTG
GAACAAAGCTAATGATCTTCTCACACAAGAACACGCAGCTTTTCTCAATGATCCTACCA
TCTCATGTTAGATCCACCTCCCGAAACCTTAATTCACCTGGACGAAGACGAAGAGTACGA
TGAAGACATGGATGCGATGAAGGAGATGCAGTACATGATCGCCGTCATGCAGCCCGTAGA
CATCGACCTGCCACGGTCCCTAAGCCGAACCGCGTAACGTAAGGATAAGCGACGATCC
TCAGACGGTGGTTGCTCGTCCGCGTCCGGAAAGGATCAGCGAGAAGATCCGAATTTCTAA
GAGGATCGTGCTGGTGGTGCAGAGATGGACACAGCTTCCATGCTCGACGAAGCCATACG
TTACACCAAGTTCTTGAAACGGCAGGTGAGGATTCTTCAGCCTCACTCTCAGATTGGAGC
TCCTATGGCTAACCCCTCTTACCTTTGTTATTACCACAACCTCCCAACCCGTATGAACCTAC
ACAGAAGCTCGCTAGCTAGACATTTGGTGTCTCCTCTCAACCTTT

>G779 Amino Acid Sequence (domain in AA coordinates: 126-182)
MKMENGMKKKGVCDSCVSSKSRSNHSPKRSMEPQPHLLMDWNKANDLLTQEHAAFLN
DPHHLMLDPPPETLIHLDEDEEDMDAMKEMQYMIAMQPVVIDPATVPKPNRRNVRI
SDDPQTVVARRRRRERISEKIRILKRIVPGGAKMDTASMLDEAIRYTKFLKRQVRILQPHS
QIGAPMANPSYLCYHNSQP*

>G962 (148..1392)

CGTCGACTCTCTACTCAACACCACTCAATTTTCATCTCTCTTTTCCCTTCCATTGTTAGT
ATAAAAACCAAGCAAACCTTAATCACTTTTCATCATCATATATCACCTTAATCCACATG
CATACACATATCTAGTCTTTTGTATATATGGCAATTGTATCCTCCACAACAAGCATCATT
CCCATGAGTAACCAAGTCAACAATAACGAAAAAGGTATAGAAGACAATGATCATAGAGGC
GGCCAAGAGAGTCATGTCCAAAATGAAGATGAAGCTGATGATCATGATCATGACATGGTC
ATGCCCCGATTTAGATTCCATCTACCGAAGAAGAACTCATAGAGTTTTACCTTCGCCGA
AAAGTTGAAGGCAACGCTTTAATGTAGAAGTCACTTTCTCGATCTTTATCGCTAT

GATCCTTGGGAACCTTCCTGCTATGGCGGCGATAGGAGAGAAAGAGTGGTACTTCTATGTG
CCAAGAGATCGGAAATATAGAAATGGAGATAGACCGAACCGAGTAACGACTTCAGGATAT
TGGAAAGCCACCGGAGCTGATAGGATGATCAGATCGGAGACTTCTCGGCCTATCGGATTA
AAGAAAACCCTAGTTTTCTACTCTGGTAAAGCCCCCTAAAGGCACTCGTACTAGTTGGATC
ATGAACGAGTATCGTCTTCCGCACCATGAAACCGAGAAGTACCAAAAGGCTGAAATATCA
TTGTGCCGAGTGTAACAAAGGCCAGGAGTAGAAGATCATCCATCGGTACCACGTTCTCTC
TCCACAAGACATCATAACCATAACTCATCGACATCATCCCGTTTAGCCTTAAGACAACAA
CAACACCATTATCCTCCTCTAATCATTCCGACAACAACCTTAACAACAACAACAACATC
AACAACTCTCGAGAAGCTCTCCACCGAATATTCCGGCGACGGCAGCACAACAACAACGACC
ACAAACAGTAACCTTGACGTTACCATTGCTCTAGCCAATCAAAACATATATCGTCCAATG
CCTTACGACACAAGCAACAACACATTGATAGTCTCTACGAGAAATCATCAAGACGATGAT
GAAACTGCCATTGTTGACGATCTTCAAAGACTAGTTAACTACCAAATATCAGATGGAGGT
AACATCAATCACCAATACTTTCAAATTGCTCAACAGTTTCATCATACTCAACAACAAAAT
GCTAACGCAAAACGCATTACAATTGGTGGCTGCGGCGACTACAGCGACAACGCTAATGCCT
CAAACCTCAAGCGCGCTTAGCTATGAACATGATTCTGCGAGGAACGATTCCAAACAATGCT
TTGTGGGATATGTGGAATCCAATAGTACCAGATGGAAAACAGAGATCACTATACTAATATT
CCTTTTAAGTAATTTAATTAGATCATGATTATTATCCATGACAATAATTAATGCTGCTTT
GCGC

>G962 Amino Acid Sequence (domain in AA coordinates: 53-175)
MAIVSSTSIIPMSNQVNNEKGIENDHRGGQESHVQNEDEADDHDMVMPGFRFHPT
EEELIEFYLRKVEGKRFNVELITFLDLRYDPWELPAMAAIGEKEWYFYVPRDRKYRNG
DRPNRVTTSGYWKATGADRMIRSETSRPIGLKTLVIFYSGKAPKGRTRTSWIMNEYRLPHH
ETEKYQKAEISLCRVYKRPVEDHPSVPRSLSTRHNNHNSSTSSRLALRQQQHSSSSNH
SDNNLNNNNNINNLEKLSTEYSGDGSTTTTTTNSNSDVITIALANQNIYRPMYDTSNNTL
IVSTRNHQDDDETAIVDDLQRLVNYQISDGGNINHQYFQIAQQFHHTQQQNANANALQLV
AAATTATTLMPQTQAALAMNMIPAGTIPNNALWDMWNPIVDPGNRDHYTNIPFK*

>G977 (46..591)
CACCAAACCTCACCTGAAACCCCTATTTCCATTTACCATTACACTAATGGCACGACCACAA
CAACGCTTTTCGAGGCGTTAGACAGAGGCATTGGGGCTCTTGGGTCTCCGAAATTTCGTCAC
CCTCTCTTGAAAACAAGAATCTGGCTAGGGACGTTTGAGACAGCGGAGGATGCAGCAAGG
GCCTACGACGAGGCGGCTAGGCTAATGTGTGGCCGAGAGCTCGTACTAATTTCCCATAC
AACCCCTAATGCCATTCTTACTTCTCTTCCAAGCTTCTATCAGCAACTCTTACCGCTAAA
CTCCACAAATGCTACATGGCTTCTCTTCAAATGACCAAGCAACGCAACACAAACGCAA
ACGCAGACCGCAAGATCACAATCCGCGGACAGTGACGGTGTGACGGCTAACGAAAGTCAT
TTGAACAGAGGAGTAACCGGAGACGAGAGATCAAGTGGGAAGATGGAAATGCCAATATG
CAACAGAATTTTAGGCCATTGGAGGAAGATCATATCGAGCAAAATGATTGAGGAGCTGCTT
CACTACGGTTCCATTGAGCTTTGCTCTGTTTACCAACTCAGACGCTGTGAGAAATGGCC
TTGTCTGTTTTAGCGTATTCTTTTCATTTTTATTTTTGTTTCCACAAAAACGGCGTCGTA
GTGATGAGAGTAGTAGTGAGAGAAGGCTAATTTCAAGACATTTTGATCTGAATTGGCCTC
TTTTGAAACACTGATTCTAGTTTTCTATAAGAGCAATCGATCATATGCTATGTTATGTATA
GTATTATAAAAAAATGTTATTTTCTGATTNAAAAAAAAAAAAAAAAAAAAAAAAA

>G977 Amino Acid Sequence (domain in AA coordinates: 5-72)
MARPPQQRFRGVRQRHWGSWVSEIRHPLKTRIWLGTFFETAEDAARAYDEAARLMCGPRAR
TNFPYNPNAPIPTSSSKLLSATLTAKLHKCYMASLQMTKQTQTQTQTARSQSADSDGVT
ANESHLNRGVTETTEIKWEDGNANMQNFRPLEEDHIEQMIIELLHYGSIELCSVLPTQT
L*

>G1063 (241..966)
GTTAAAGAAGATGGATGGGCCCACAAGTTGCTATATAAATCCTTCCACTTCTTGTGTATA
CTATTGCTTGAGTTCTGATTGGGCACAGTAGTACCATTGCCATTTCTCTCACACATACCG
TCTCTTTCTCTCATCATCAATCATCAATCATCCAAAAGAAAAACCCATAAATTTCACTT
GTAAGCTTTTACCAGTTTCTCTCCATACCCATTTTATCAGCTTCTCCATATCTTTCTCT
ATGGATTCTGACATAATGAACATGATGATGCATCAGATGGAGAAGCTTCTGAGTTTGT
AACCCCTAATCCTCTTTCTCTCTCCGACCACAACAACACTTACCCTTTTCTCTTTAAC
TCCACTCATTACCAGTCCGATCACTCAATGACCAACGAACAGGTTTCCGCTACGGTTCC
GGTTTACTCACTAACCCCTTCTCTATCTCTCCCAACACAGCTTACTCTTCCGTTTTTCTT
GACAAAAGAAACAACAGTAACAACAACAATAATGGCACGAACATGGCAGCTATGCGAGAG
ATGATCTTCCGTATCGCCGTGATGCAACCGATCCATATCGATCCCGAGGCGGTTAAGCCA

CCGAAGAGGAGGAACGTCAGGATCTCTAAAGATCCTCAAAGCGTGGCGGC'TAGGCATAGA
AGGGAGAGAATAAGCGAGAGGATTTCGGATTTTGGCAACGGCTTGTTCCTGGTGGGACGAAG
ATGGATACAGCTTCGATGCTCGATGAAGCAATTCATTATGTGAAGTTTTTAAAGAAACAG
GTGCAGTCTCTGGAGGAGCAGGCGGTGGTTACTGGCGGAGGGGGAGGAGGAGGAGGAAGG
GTTTTGATCGGTGGAGGTGGAATGACGGCGGCGAGTGGTGGTGGTGGCGGGGGAGTG
GTTATGAAAGGGTGTGGAACAGTGGGACTCATCAGATGGTGGGCAATGCACAGATTCTT
AGATGATGATGATTTTAAATTTTATTATTATTAATGTTGGAGAAAAAGAGAAAAA
TGATTCTGGAGAGGGAAGCCAAGTAATTTATGTGAGAGTCTTTAATTTAACTTTATTTTC
TTGTTTAGATAATGTGTAATGATGGTTTTTAAAGCCAAAGACTCTCCATGGTTGTGGAG
CGAGTTTG

>G1063 Amino Acid Sequence (domain in aa coordinates: 131-182)

MDSDIMNMMMHQMEKLPEFCNPNSFFSPDHNNTPFLFNSTHYQSDHSMTNPEGFRYGS
GLLTNPSSISPNTAYSSVFLDKRNNNNNNNGTNMAAMREMIFRIAVMQPIHIDPEAVKP
PKRRNVRIKDPQSVAAHRHRRERISERIRILQRLVPGGTKMDTASMLDEAIHYVKFLKKQ
VQSLLEEQAVVTGGGGGGGGRVLIGGGGMTAASGGGGGGGVVMKCGTVGTHQMVNAQIL
R*

>G1140 (67..729)

ATCCAAGATCCTCCAACCTACAGAAAGGCAGATTCAAGAACAGTAGTGAAGGAGAGATCT
GGTAAATGGCGAGAGAGAAGATAAGGATAAAGAAGATTGATAACATAACAGCGAGACAA
GTTACTTTCTCAAAGAGAAGAAGAGGAATCTTCAAGAAAGCCGATGAACCTTTCACTTCTT
TGCGATGCTGATGTTGCTCTCATCATCTTCTCTGCCACCGGAAAGCTCTTCGAGTTCTCC
AGCTCAAGAATGAGAGACATATTGGGAAGGTATAGTCTTCATGCAAGTAACATCAACAAA
TTGATGGATCCACCTTCTACTCATCTCCGGCTTGAGAATTGTAACCTCTCCAGACTAAGT
AAGGAAGTCGAAGACAAAACCAAGCAGCTACGGAAGTGAAGAGGAGAGGATCTTGATGGA
TTGAACTTAGAAGAGTTGACGCGCTGGAGAACTACTTGAATCCGGACTTAGCCGTGTG
TCTGAAAAGAAGGGCGAGTGTGTGATGAGCCAAATTTTCTCACTTGAGAAACGGGGATCG
GAATTGGTGGATGAGAATAAGAGACTGAGGGATAAACTAGAGACGTTGGAAGGGCAAAA
CTGACGACGCTTAAAGAGGCTTTGGAGACAGAGTCCGTGACCACAAATGTGTCAAGCTAC
GACAGTGAAGTCCCTTTGAGGATGACTCCGACACTTCCCTGAAGCTTGGGCTTCCATCT
TGGGAATGAATCTGAGAGAGAGAAAGATCCAGCAGAGTTGACTTCGATGGAAGCCCACAA
ATATTAAGTCTACCTTTTCCCTTTCTTTTCTTGAATAAGTGTTGAAAAAGAATTGAGAT
GGGAAGGATGAATTCTCATTGCATTGCAGAGAAGCAAGTTTCAGATATTGTACGTGTTAT
TGGGTCTTTATAACTATTTTCTCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>G1140 Amino Acid Sequence (conserved domain in AA coordinates:2-57)

MAREKIRIKKIDNITARQVTFKRRRGIFKKADELVLCDADVALIIFSATGKLFEFSSS
RMRDILGRYSLHASNINKLMDPPSTHLRLCNCNLSRLSKEVEDKTKQLRKLRGEDLDGLN
LEELQRLKLLSGLSRVSEKKGECVMSQIFSLEKRGSELVDENKRLRDKLETLEAKLT
TLKEALETESVTTNVSSYDSGTPLEDDSDTSLKLGLPSWE*

>G1425 (43..1005)

ACTCTCTCAAACCATAAAAAATATTCTCCGATCATCATTTTAAATGGAGAGTACAGATTCT
TCCGGTGGTCTCCGCGCGCGCAACCAACCTCCCTCCAGGATTCCGGTTTCATCCAACA
GACGAAGAACTTGTAATTCATTACCTCAAACGCAAAGCAGATTCTGTTCCCTTTACAGTC
GCGATCATCGCCGACGTTGATCTTTACAAATTTGATCCATGGGAAC'TCCCGCGAAAGCT
TCGTTTGGAGAACAAGAATGGTATTTTTTTCAGTCCAAGAGATCGGAAATATCCCAACGGA
GCTAGACCTAACCGAGCTGCGACTTCCGGTTATTGGAAAGCGACTGGTACAGATAAACCG
GTGATTTCAACCGCGCGGTGGTGGTAGTAAAAAAGTGGGAGTTAAAAAGGCTCTAGTGTTT
TACAGTGGTAAACCACCAAAAGGAGTTAAATCAGATTGGATTATGCATGAATATCGGTTA
ACTGATAATAAACCTACTCACATTTGTGACTTCGGCAACAAGAAAACTCTCTCAGGCTT
GATGATTGGGTGTTGTGTCGTATCTACAAGAAAAACAATAGTACAGCATCTAGACATCAT
CATCATCTTCATCATATTATCATCTAGATAATGATCATCATCGTCATGATATGATGATTGAT
GATGATCGATTCCGTCATGTTCTCTCGGTCTTCACTTCCCGGCGATTTTTTCTGACAAT
AATGATCCGACGCTATATATGATGGTGGCGGCGCGGATACGGAGGTGGAAGTTACTCG
ATGAATCATTTGTTTCGCATCTGGATCAAAGCAGGAGCAGTTGTTTCCACCGGTGATGATG
ATGACTAGTCTAAATCAAGATTCCGGTATTGGATCGTCGTCTGTCACCTAGCAAGAGATTT
AACGGCGGCGCGGTGGAGATTGTTGCACTTCTATGGCGCGGACGCGGTTAATGCAGAAC
CAAGGTGGGATTTACCAATTGCCTGGTTTGAATTGGTATTCTTGAAAAACAATTTACGATG
AAGAATTTTTAAATTTGTGTATATATACGGTTTGAGTGATTAGGGGGCATTGGGGGA

TTTATTTACGGTTGATTATTATTGTAGTGTATAGAACTAAGGAGATTAAATTAAATAGA
TTGGAGGAAAAAAAAAAAAAAAAA

>G1425 Amino Acid Sequence (domain in AA coordinates: 20-173)
MESTDSSGGPPPPQPNLPPGFRHPTDEELVIHYLKRKADSVPLPVAIADVDLYKFDPW
ELPAKASFGEQEWYFFSPDRKYPNGARPNRAATSGYWKATGTDKPVISTGGGSSKKVGV
KKALVFYSKPPKGVKSDWIMHEYRLTDNKPETHICDFGNKKNSLRLLDDWVLCRIYKKNNS
TASRHHHHLHHIHLNDNDHHRHDMIDDDRFRHVPPGLHFPALFSDNNDPTAIYDGGGGGY
GGGSYSNMNHCFASSGSKQEQLFPPVMMMTSLNQDSGIGSSSSPSKRFNGGGVGD CSTSMAA
TPLMQNQGGIYQLPGLNWYS*

>G1449 (105..581)
TAGACAGAGAGAAATAGAGAGAGAGACATGAAGAGCACTCTCAATAGAGAAG
AGAAGGAAGCATGAAGCTAGCTCTGCAGCTTCAAGGTCTCATTATGGAGGTCTCTAACT
CTTGTTCTTCATTTTCTTCATCCTCTGTCGACAGTACTAAACCTTCTCCTTCTGAATCTT
CTGTTAATCTCTCCCTTAGTCTCACATTTCTTCTACTTCTCCACAAAGAGAAGCAAGAC
AAGATTGGCCACCGATAAAGTCTAGATTAAGAGATACACTAAAGGGTCGTCGTCTTCTTC
GTCGTGGTGATGACACTTCTCTCTTTGTTAAGGTTTATATGGAAGGTGTTCCCATGGAA
GAAAACTCGACCTTTGCGTATTCTCAGGCTACGAGAGTCTATTAGAAAATCTCTCTCACA
TGTTTCGATACTTCAATCATCTGCGGTAAATCGAGATCGAAAACATCATGTTTTGACATATG
AAGACAAGGATGGAGATTGGATGATGGTCGGAGATATCCATGGGATATGTTTCTTGA
CCGTGAGAAGACTAAAGATCACGAGACCGGAGAGGTATTAAGGTTGGATCGGTCAAGGC
TGTGATTGCGCAGTTACGAGACGTGTAAGATTTAGGCATTGATGAAGAGACTTGAGGCGG
GACGGAGCTATTGCTGCATATTGCAACAAAGGCCTTGAAGAAGTTGGAGAATTGATTGAT
GCATATATTTATTTATATGACACCTTTGAGTGTGTTTTTCTTATAAATAAATCACAATA
TCCAAGACTTCTCTTTAAA

>G1449 Amino Acid Sequence (domain in AA coordinates: 48-53,74-107,122-152)
MEVSNSSCSFSSSSVDSTKPSPESSVNLSSLTFPSTSPQREARQDWPIKSRLRDLTK
GRRLLRGGDDTSLFVKVMEGVPIGRKLDLCVFSGYESLLENLSHMFDTSIICGNRDRKH
HVLTYEDKGDWMMVGDIPWDMFLETVRRLKITRPERY*

>G1897 (1..678)
ATGCCTTCTGAATTCAGTGAATCTCGTCGGGTTCTTAAGATTCCCCACGGCCAAGGAGGA
TCTGTTGCGATTCCGACGGATCAACAAGAGCAGCTTTCTTGTCTCGCTGTGAATCAACC
AACACCAAGTTCTGTACTACAACAACACTCAACTTCTCACAACCTCGTCATTTCTGCAAG
TCTTGTGCGCGTTACTGGACTCATGGAGGTACTCTCCGTGACATTTCCCGTCGGTGGTGT
TCCCGTAAAAGCTCAAAACGTTCCCGGACTTATTCCTCTGCCGCTACCACCTCCGTTGTC
GGAAGCCGGAACCTTCCCTTACAAGCTACGCTGTTCTTTTCCCTCAGTCGTCTTCCAAC
GGCGGTATCACGACGGCGAAGGGAAGTGCTTCGTGCTTCTATGGCGGTTTCAGCTCTTTG
ATCAACTACAACCGCGCGGTGAGCAGAAATGGGCTGGTGGCGGGTTAATGGGCCAGAT
GCTTTTGGTCTTGGGCTTGGTTCAGGGTCTGATTATGAGGACGTCAGATGGGCAAGGA
ATAACGGTCTGGCGGTTTCAAGTGGCGCTACTGATGCTGCAACTACTACAAGCCACATT
GCTCAAATACCCGCCACGTGGCAGTTTGAAGGTCAAGAGAGCAAAGTCGGGTTCTGTCT
GGAGACTACGTAGCGTGA

>G1897 Amino Acid Sequence (domain in AA coordinates:34-62)
MPSEFSESRRVPKIPHGQGSVAIPTDQEQQLSCPRCESTNTKFCYYNNYNFSQPRHFCK
SCRRYWTHGGTLRDI PVGGVSRKSSKRSRTYSSAATTSVVGSRNFPLOATPVLFPQSSN
GGITTAKGSASSFYGGFSSLINYNAAVSRNGPGGGFNGPDAFGLGLGHGSYYEDVRYGQG
ITVWPFSSGATDAATTSHIAQIPATWQFEGQESKVG FVSGDYVA*

>G2143 (89..784)
TCTTCTTCTCTCTCCATACCTTATCTCACCAGCTTCTCCATATCTCTCAAAGAAAAACA
AACCTTATAAATCCACAAAAAGGAGGATGGATAACTCCGACATTCTAATGAACATGAT
GATGCAGCAGATGGAGAAGCTTCTGAACACTTCTCTAACTCAAACCTAACCCTAATCC
CCATAACATTATGATGCTTTCTGAATCCAACACCCACCCGTTCTTCTTCAACCCCACTCA
TTCTCATCTCCCATTTGACCAAACCATGCCTCACCACCAACCCGTTTAAATTTCCGGTA
CGCCCCCTCCCGTCATCATCTCTCCCGGAGAAGAGAGGAGGCTGCAGCGACAACGCCAA
CATGGCGGCGATGAGAGAGATGATCTTTCGAATAGCCGTGATGCAGCCTATACATATTGA
TCCGGAATCCGTAAGACCACCAAAGAGAAAGAGAGGATCTCTAAGGATCCACAGAG
CGTGGCAGCTCGGCATCGAAGGGAGAGGATAAGCGAGCGGATTCCGATTCTTCAGCGGCT
TGTTCCCGGTGGGACTAAGATGGATACGGCGTCGATGCTCGATGAGGCTATCCATTACGT

TAAGTTTCTCAAGAAGCAAGTGCAGTCGCTGGAGGAACATGCGGTGGTTAACGGCGGAGG
 AATGACGGCGGTGGCCGGAGGAGCACTTGCGGGTACTGTTGGTGGAGGATATGGAGGAAA
 AGGGTGTGGCATTATGCGGTCTGATCATCACCAGATGCTTGGAAATGCACAGATTCTTAG
 ATGATGATGATGTTGATTTTTAAATATATATCATATGTTTATTAATATGACGGGAAAAAA
 TATTATCGAGGGAGTTGAATTTAGTATCATGAAACTATGAGAGCATTTTTTTAAATGTT
 TTTATCTTTCCGGGTTTCGATAATGTTTGGGATGGTTAATTAACAATTTAAAAGTCAGAC
 AACTTGGTGTAAAGACTAAAGAATAAGCATAGTTTATCAATTTATCATTACTAAATGAA
 ATAG

>G2143 Amino Acid Sequence (domain in aa coordinates: 128-179)

MDNSDILMNMQQMEKLPEHFSNSNPNNPHNIMMLSESNTHPFFNPTHSHLPFDQTM
 PHHQPLNFRYAPSPSSSLPEKRGCSNANMAAMREMI FRIAMQPIHIDPESVKPPKR
 KNVRISKDPQSVAAHRHRERISERIRILQRLVPGGTMKTASMLDEAIHYVKFLKKQVQS
 LEEHAVVNGGGMATAVAGGALAGTVGGGYGGKGGIMRSDHQLGNAQILR*

>G2535 (1..1005)

ATGAACATATCAGTAAACGGACAGTCACAAGTACCTCCTGGCTTTAGGTTTCACCCAACC
 GAGGAAGAGCTCTTGAAGTATTACCTCCGCAAGAAAATCTTAACATCAAGATCGATCTC
 GATGTTATCTCTGACATTGATCTCAACAAGCTCGAGCCTTGGGATATTCAAGAGATGTGT
 AAGATTGGAACGACGCCGCAAAACGATTGGTACTTTTATAGCCATAAGGACAAGAAGTAT
 CCCACCGGGACTAGAACCAACAGAGCCACCACGGTCGGATTTTGGAAAGCGACGGGACGT
 GACAAGACCATATATACCAATGGTGATAGAAATCGGGATCGGAAAGACGCTTGTCTTCTAC
 AAAGGTCGAGCCCCATGGTTCAGAAATCCGATTGGATCATGCACGAATATAGACTCGAC
 GAGAGTGATTAATCTCTCTCGTGTGGCGATCATGACGTCAACGTAGAAACGTGTGATGTC
 ATAGGAAGTGACGAAGGATGGGTGGTGTCTGCTGTTTTCAAGAAAAATAACCTTTGCAAA
 AACATGATTAGTAGTAGCCCGCGAGTTCGGTGAAAACGCCGTCGTTCAATGAGGAGACT
 ATCGAGCAACTTCTCGAAGTTATGGGGCAATCTTGTAAGGAGAGATAGTTTTAGACCCT
 TTCTTAAAACTCCCTAACCTCGAATGCCATAACAACACCACCATCACGAGTTATCAGTGG
 TTAATCGACGACCAAGTCAACAACCTGCCACGTCAGCAAAGTTATGGATCCCAGCTTCATC
 ACTAGCTGGGCCGCTTTGGATCGGCTCGTTGCCTCACAGTTAAATGGGCCCCAATCGTAT
 TCAATACCAGCCGTTAATGAGACTTCACAATCACCGTATCATGGACTGAACCGGTCCGGT
 TGTAAATCCGTTTAAAGACGAGATTACTATATACCGGAGATTGATTTATGGAACGAGGCA
 GATTTTCGCGAGAACGACATGCCACTTGTGTAACGGTAGTGGATAA

>G2535 Amino Acid Sequence (conserved domain in AA coordinates: 11-114)

MNISVNGSQVPPGFRFHPTEBELLKYLYLRKKISNIKIDLDVIPDIDLNLKLEPWDIQEMC
 KIGTTPQNDWYFYSHKDKICYPTGTRTNRATTVGFWKATGRDKTIYTNDRIGMRKTLVFI
 KGRAPHGQKSDWIMHEYRLDESVLISSCGDHVDNVETCDVIGSDEGWVVCRVFKNNLCK
 NMISSPASSVKTPSFNEETIEQLLEVMGQSKGEIVLDPFLKLPNLECHNNTTITSYQW
 LIDDQVNNCHVSKVMDPSFITSWAALDRLVASQLNGPNSYSIPAVNETSQSPYHGLNRS
 CNTGLTPDYIPEIDLWNEADFARTTCHLLNGSG*

>G2557 (94..1215)

TCGACTTCCTGTGAACATCATCTGTTTGTCTCTTCTTCCGGTTTCACTTTTTTCATGTCCT
 GCCGTTATTACAACGAGGATTGTGTTTGATCCGATGGAAGGATTGGAATCTGTGTACGCT
 CAAGCTATGTATGGAATGACACGAGAGAGCAAAATCATGGAGCATCAAGGATCAGATTG
 ATTTGGGGAGGAAATGAGCTAATGGCTCGAGAACTCTGTTCTTCTTCTTCTTATCACCAC
 CAACTCATTAAATCCGAATCTTAGCAGCTGTTTCATGTCTGATCTTGGAGTCTTAGGTGAG
 ATTCAACAGCAGCAACATGTTGGCAACAGAGCTAGCTCGATAGATCCATCATCACTCGAT
 TGTGTTGTTATCTGCGACGTGCAATAGCAACAACACCTCGACGGAGGACGATGAAGGAATA
 TCTGTGCTTTTCTCAGATTGTCAGACTCTTTGGAGCTTTGGTGGAGTCTCATCTGCAGAG
 TCTGAGAACAGAGAGATCACTACTGAGACGACAACAACGATAAAGCCTAAGCCTTTGAAG
 AGAAACAGAGGAGGAGATGGAGGAACTACTGAGACTACAACAACAACAAAACCTAAG
 TCTTTGAAGAGAAACAGAGGAGACGAGACAGGAAGTCACTTAGTCTTGTTCATCCTCAA
 GATGATTCGGAGAAAGGAGGTTTCAAGCTTATATACGATGAGAATCAATCGAAATCAAAG
 AAACCAAGAACAGAGAAAGAACGAGGCGGTTCTTCAACATTAGTTTCCAACATTCAACT
 TGTGTTGCTGACAATGTCGAGCCCGATGCTGAGGCGATTGCACAAATGAAGGAGATGATA
 TACAGAGCGGCTGCATTTAGACCGGTGAATTTCCGGTTAGAGATTGTGGAGAAGCCTAAG
 AGGAAGAACGTCAAGATATCGACGGATCCTCAAACGGTTGAGCGAGACAGAGAAGGGAG
 AGGATAAGTGAGAAGATTAGGGTTTTACAAACATTGGTTCCAGGTGGGACGAAGATGGAT
 ACTGCATCAATGCTTGATGAAGCTGCTAATTATCTCAAGTTTCCTTAGAGCACAAAGTAAAA

GCTTTTAGAAAACCTTGAGACCCAAGCTTGACCAAACCAATCTCTCTTTCTCTTCTGCTCCT
ACATCGTTTCCATTATTCCACCCATCTTTTCTTCCATTGCAAAATCCTAATCAAATCCAT
CATCCAGAGTGTGACAGATTATAAACTTTTGTAGTTTCATCATCATCAACAGAATCATGG
CGTCTTGATTGTTTTAGCAGTTCTCAAGAAAGGCAACTTCTGTGACAAGGGTGGTGTGCG
GCAGTGTGTTTACACTTTCCAGTCTTTGTTTGCATTTCTTTTATATAAAAGTTTGTAT
TTTATATAGAATCTGTGGAATTCGAGGGTTGAAATATTGTGAAAAACAGAGCCGCAAGAG
GTTAATTACAGTCTCTGCAATATTTCAACCTTTTATTACTTTATTAGAGTAAAGATAGC
GT

>G2557 Amino Acid Sequence (domain in aa coordinates: 278-328)

MEGLESVYAQAMYGMTRESKIMEHQGSDLIWGGNELMARELCSSSSYHHQLINPNLSSCF
MSDLGVLGELIQQQHVGNRRASSIDPSSLDCLLSATSNNSNTSTEDDEGISVLFSDCQTLW
SFGGVSSAESENREITTETTTTIKPKPLKRNRRGGDGGTTETTTTTPKPSLKRNRRGDETG
SHFSLVHPQDDSEKGGFKLIYDENQSKSKPRTEKERGGSSNISFQHSTCLSDNVEPDAE
AIAQMKEMIYRAAAFREPVNFGLEIVEKPKRKVKIISTDPQTVAARQRRERISEKIRVLQT
LVPGGTKMDTASMLDEAANYLKFLRAQVKALENLRPKLDQTNLSFSSAPTSFPLFHPSPFL
PLQNPNIHHPEC*

>G259 (52..786)

GAGATCTTCTACTACTTGTGTTTTCTCAAGAATAATAATTTTCGTTTTATATATGGAAGAT
GCTGGTGAAACATTTACGGTGTAACGATAACGTTAACGACGAGGAGCGTTTGCCATTGGAG
TTTATGATCGGAACTCAACATCCACGGCGGAGCTACAGCCGCTCCACCGTTCTTGGA
AAGACATACAAAGTGGTGGAGGATCCGACGACGCGGGGTTATATCTTGGAAACGAATAC
GGAAGTGGTTTTCGTGTGCGGAGCCGCGCAGAAATTCGCTAGAGATCTGTTACCAACACTT
TTCAAGCATTGCAACTTCTCTAGCTTCGTTCCGCGAGCTCAATACTTACGGTTTTCGAAAA
GTAACGACGATAAGATGGGAATTTAGTAATGAGATGTTTCGAAAGGGGCAAAGAGAGCTT
ATGAGCAATATCCGAAGAAGGAAGCAACATTTGGTCACACAACAAGTCTAATCACCAG
GTTGTACCAACAACAACGATGGTGAATCAAGAAGGTCATCAACGGATTGGGATTGATCAT
CACCATGAGGATCAACAGTCTTCCGCCACTTCATCCTCTTTTCGTATACACTGCATTACTC
GACGAAAAACAAATGCTTGAAGAATGAAAAACGAGTTATTAAGCTGCGAACTTGGGAAAACC
AAGAAGAAATGCAAGCAGCTTATGGAGTTGGTGGAGAGATACAGAGGAGAAGACGAAGAT
GCAACTGATGAAAGTGATGATGAAGAAGATGAAGGGCTTAAGTTGTTCCGGAGTAAACTT
GAATGAAACTAGATTGCTAGATTGATATTCTGTAATATACCAGTTTCTTCATATTCTTAGA
AGTTTTGCATAACTATATATAGTACTCTTTTAAGACATGCAAGATCAGAACATATG

>G259 Amino Acid Sequence (domain in AA coordinates: 27-131)

MEDAGEHLRCNDNVNDEERLPLEFMIGNSTSTAELOPPPPFLVKYTKVVEDPTTDGVISW
NEYGTGFVWQPAEFARDLLPTLFKHCNFSFVRQLNTYGRKVTITIRWEFSNEMFRKGQ
RELMSNIRRRKSQHWSHNKSNNHQVPTTTMVNQEGHQRIQIDHHHEDQSSATSSSFVYT
ALLDENKCLKNENELLSCELGKTKKKCKQLMELVERYRGEDEDATDESDDDEDEGLKLF
VKLE*

>G353 (82..570)

ACCAAACCTCAAAAAACACAAACCACAAGAGGATCATTTTCATTTTTTATTGTTTCGTTTTA
ATCATCATCATCAGAAGAAAAATGGTTGCGATATCGGAGATCAAGTCGACGGTGGATGTC
ACGGCGGCGCAATTGTTTGATGCTTTTATCTAGAGTTGGACAAGAAAACGTTGACGGTGGC
GATCAAAAACCGGTTTTTACATGTAAAACGTGTTTGAAGCAGTTTCATTTCGTTCCAAGCC
TTAGGAGGTCACCGTGCGAGTCACAAGAAGCCTAACAACGACGCTTTGTCGTCTGGATTG
ATGAAGAAGGTGAAAACGTCGTCGCATCCTTGTCCCATATGTGGAGTGGAGTTTCCGATG
GGACAAGCTTTGGGAGGACACATGAGGAGACACAGGAACGAGAGTGGGGCTGCTGGTGGC
CGCTTGGTTACACGCGCTTTGTTGCCGAGCCACGGTGACTACGTTGAAGAAATCTAGC
AGTGGGAAGAGAGTGGCTTGTGATCTGAGTCTAGGGATGGTGGACAATTTGAATCTC
AAGTTGGAGCTTGAAGAACAGTTTATTGATTTTATTTATTTTCCTTAAATTTTCTGAAT
ATATTTGTTTCTCTCATTCTTTGAATTTTCTTAAATATTCTAGATTATACATACATCCGC
AGATTTAGGAACTTTCATAGAGTGAATCTTTTCTTTCTGTAAAAATATATTTTACTTG
TAGCAA

>G353 Amino Acid Sequence (domain in aa coordinates: 41-61, 84-104)

MVAISEIKSTVDVTAANCLMLLSRVGQENVDDGQKRVFTCKTCLKQPHSFQALGGHRAS
HKKPNNDALSSGLMKKVKTSHPICGVEFPMGQALGHHMRRHRNESGAAGGALVTRAL
LPEPTVTTLKSSSGKRVACLDLSLGMVDNLNLKLELGRTVY*

>G354 (27..533)

CCTAGAAGTCACTAAGTCGATTCAAAATGGTTGCGAGAAGTGAGGAAATTGTGATAGTGG
AAGAAGATACGACTGCGAAATGTTTGATGTTGTTATCAAGAGTCGGAGAATGCGGCGGCG
GCTGCGGGGAGATGAACGTGTTTTCCGATGCAAGACTTGTCTTAAAGAGTTCTCATCGT
TTCAAGCTTTGGGAGGTCATCGTGCAAGCCACAAGAACTTATCAACAGTGACAATCCAT
CACTTCTTGGATCCTTGTCCAACAAGAAAATAAAACGTCTCATCCTTGTCCGATATGTG
GAGTGAAGTTTCCGATGGGACAAGCTCTTGGTGGTCACATGAGGAGACATAGGAACGAGA
AAGTCTCAGGCTCGTTGGTTACACGTCTTTTTCTACCGGAGACGACGACGGTGACGGCTT
TGAAGAAATTTAGTAGTGGGAAGAGAGTGGCTTGTGTTGGATTGGACTTAGATTTCGATGG
AGAGTTTGGTCAATTGGAAGTTGGAGTTGGGAAGAACGATTCTTGGAGTTAAGTTTGTG
GGTTGTATACAGTTTACATGATTTTGTATCTTTGTTGATCCAATTATCGTACCGATCG
ATGTGAATATATTTTGTATACAATAAAA

>G354 Amino Acid Sequence (domain in aa coordinates: 42-62, 88-109)

MVARSEEIVIVEEDTTAKCLMLLSRVGECGGGCGDERVFRCKTCLKEFSSFQALGGHRA
SHKKLINSNDNPSLLGSLSNKKTSHPCPICGVKPFMQALGGHMRRHRNEKVSGSLVTR
SFLPETTTVTALKKFSSGKRVACLDDLDSMESLVNWKLELGRITISWS*

>G638 (86..1861)

GAATTAAGGTTTAACTTTTACCTTTTTTCCCTTCACTATCGATAATTGATCTTCTCT
TTCGGCTGAATATAAATCTGAAAAAATGGATCAAGATCAGCATCCTCAGTACGGTATACC
GGAGCTCCGGCAGCTCATGAAAGGCGGAGGAAGGACGACTACTACAACACCGTCTACTTC
TTCTCATTTTCCCTCTGATTCTTTCGGTTTAACTTGTCTCCGGTGCAGCCACCGCCACA
CCGCTTTCATCAGTTCACTACTGATCAAGATATGGGTTTCTTGGCACGTGGCATAACATGG
ATTGGGTGGAGGTTCTTCAACGGCTGGAATAACAGTAACCTTAAACCGCAGTACTAGTGG
TGGAGGAGTTGGGTTTAGTGGGTTTCTTGACGGTGGTGGTTTCGGCAGCGGAGTAGGAGG
AGACGGTGGAGGAAGTGAAGGTGGCCGAGACAAGAAACCCTAACTCTGTTGGAAATTAG
ATCTCGTCTTGTATCATAAATTCAAAGAAGCTAATCATAAAGGACCTCTTTGGGATGAAGT
TTCTAGGATTATGTCCGAGGAACATGGATACCAAAGGAGTGGGAAGAAATGCAGAGAGAA
GTTTGAGAATCTGTACAAATACTATAGTAAGACTAAAGAAGCGGAAGCCGGAAGACAAGA
CGGAAACATCACAGATTTTTCCGCCAGCTCCAAGCGCTATACGGGGATTCTAATAACTT
GGTTTCTTGTCCCAATCATAACACCGAGTTTCATGAGCAGTGCTCTTCATGGTTTCCATAC
TCAAACCCATGAACGTTGCTACAACAACGTCACATCCATAACGTTGATAGTGTTC
TGGTTTTTCATCAAAGCCTTAGTCTTCTAACAATACTAACAACCTCCCGAGCTTGGAGTAT
GACTTCTCTTTCGGAAGGGAATGATTCTAGTAGTAGAAGGAAAAAGAGGAGTTGGAAGC
GAAGATAAAGGAGTTTATTGATACGAACATGAAAGGTTGATAGAGAGGCAAGATGTTTG
GCTTGAGAAGTTGACAAAGGTTATTGAAGACAAAGAGGAACACGGATGATGAAGAAGA
GGAATGGAGGAAGATTGAAGCTGCAAGGATTGATAAAGAGCATTTGTTTTGGGCTAAAGA
GAGGGCGAGGATGGAAGCTAGGGATGTTGCGGTGATTGAGGCATTGCAATACTTGACAGG
AAAGCCATTGATAAAGCCGCTGTGTTTATCCCGGAAGAGAGGACAAATGGTAATAATGA
GATCCGAAACAATAGTGAGACACGAATGAGAATGGAAGCGATCAAACGATGACTAACAA
TGTTTGTGTTAAAGGAAGTAGTAGCTGCTGGGGTGAGCAAGAGATTTTAAAGCTTATGGA
GATAAGAACCAGCATGGACTCGACCTTTCAAGAGATATTAGGAGGGTGCTCGGATGAGTT
TCTATGGGAGGAAATCGCAGCGAAGTTGATTCAAGTATAGGGTTTATGATCAGAGAAGTGCTT
ATTATGCAAGGAAAAGTGGGAATGGATAAGCAATGGAATGAGGAAAGAAAAGAAGCAAAT
CAACAAGAAAAGAAAGGATAATTTCGTCCAGCTGCGGCGTGTACTACCCGAGAAACGAAGA
AAATCCAATCTACAATAATCGAGAAAGTGGATATAATGATAATGATCCGCATCAAATCAA
CGAACAAGGCAATGTAGGTTCTTCAACATCAAACGCAACGCAACGCAACGTAACCAC
TGGAAATCCGAGCGGTGCAATGGCTGCTAGTACAACTGCTTCCCGTTCTTCATGGGAGA
TGGAGATCAGAATTTGTGGGAGAGTTATGGTTTGAGGCTCAGTAAAGAAGAGAATCAGTA
AGTAATTTCTCTTAATGAAGAAGAAGAAGGTAATCATGTGGTTAACTAATTCTTTTGTG
TAGCTATATATGAGATAAACCTTGACTTAGCTATTATATGTCACATGCTGCTTAGAATTA
AGAAATATTTGTTGGGGCTTAACGAATTATATATCAGCATATATAAGATGAGAGTCTAAG
AATTATATCAAATTAGGCTTTAACCAACGTACGATTATATATATGTTTTCATGTATTTA
TTCTGTAAGACTTTTTAATATCAATCTTTCTCTAAA

>G638 Amino Acid Sequence (domain in AA coordinates: 119-206)

MDQDQHPQYGIPELRQLMKGGGRTTTTTPTSSHPSPDFGFNLAPVQPPPHRLHQFTTD
QDMGFLPRGIHGLGGSSSTAGNNSNLNASTSGGGVGFSGFLDGGGFGSGVGGDGGGTGRW
PRQETLTLLEIRSLDHKFKFKEANHKGPLWDEVSRIMSEEHGYQSRGKKCREKFENLYKYY
SKTKEGEAGRQDGKHHFRQLQALYGDSSNNLVSCPNHNTQFMSSALHGFHTQNPMMNVAT

TTSNIHNVDSVHGFHQSLSLSNNSSELELMTSSSEGNDSRRRKRKRSWKAKIKEFIDT
NMKRLIERQDVWLEKLTQVIEDKEEQRMKEEWRKIEAARIDKEHLFWAKERARMEARD
VAVIEALQYLTKGKPLIKPLCSSPEERTNGNNEIRNNSETQNGSDQTMNNVCVKGSSS
CWGEQEILKLMEIRTSMDSTFQEILGGCSDEFLWEBIAAKLIQLGFDQRSALLCKEKWEW
ISNGMRKEKKQINKKRKDNSSSCGVYYPNEENPIYNNRESGYNDNDPHQINEQGNVGSS
TSNANANANVTGNPSGAMAASSTNCFPFMFGDGDQNLWESYGLRLSKEENQ*

>G869 (428..1402)

AGGAACAGTGAAAGGTTCCGGTTTTTTGGGTTTCGATCTGATAATCAACAAGAAAAAGGG
TTTGATTATAGTCGGCTGGGTTTGAATCGACTGTGATTTGTCTTTGATTTCATATCTCTT
CTCCGATTTTCATCATCATCTTCCCCATCATCGTCGTCTTTGAAATCTGTCTTCTCAACG
CTCTTCACTTCTGCTGTAATAAGCAGAGGCTTGTCTGGAGACTCCTTCTCTTCCATGC
GCTTAAGACCCAAAGGACTTGTTCAGTGTGAAGTCTTTGGGGTTTTTCACATAAAGC
AGCAAAAGTTTTCTTTTTCATAGTTTCGCTGAGAGTTTGTAGTTTGTATACCAAAAAGT
TTTGACCTTTTAGAGTGATTTTTTGTCTTTCTGTTTTCTGGGTATTTTTGAGGAGTGGG
TTTAACAATGGTTGCGATTAGAAAGGAACAGTCTTTGAGTGGTGTAGTAGCGAGATTAA
GAAGAGAGCTAAGAGAAACACTCTATCGTCCCTTCTCAAGAAACCAACCTTTGAGGAA
AGTCCGTATTATTGTGAATGATCCTTATGCTACTGATGATTCTCTAGTGATGAGGAAGA
GCTTAAGGTTCTTAAGCAAGGAAAATGAAACGTATCGTTCGTGAGATTAACTTTCCTTC
TATGGAAGTTTCTGAACAGCCTTCTGAGAGTTCTTCTCAGGACAGTACTAAAAGTATGG
CAAGATAGCTGTGTGACGTTCTCTGCTGTTCTTAGGAAGAAGCCTGTTGGTGTTAGGCA
AAGGAAATGGGGGAAATGGGCTGCTGAGATTAGAGATCCTATTAAGAAAAGTAGGACTTG
GTTGGGTACTTTTGTACTCTTGAAGAAGCTGCTAAAGCTTATGATGCTAAGAAGCTTGA
GTTTGATGCTATTGTTGCTGGAATGTGTCCACTACTAAACGTGATGTTTCTTCATCTGA
GACTAGCCAATGCTCTCGTTCTTCACTGTTGTTCTCTGTTGAGCAAGATGACACTTCTGC
ATCAGCTCTCACTTGTGTCAACAACCTGATGACGCTCTGACCGTTGCTCCAAGTCTCC
AACTCCAATGTTCTGCTGGTGGAAACAAGGAAACGTTGTTTCGATTTCGACTTTACTAA
TCTACAGATCCCTGATTTTGGTTTCTTGGCAGAGGAGCAACAAGACCTAGACTTCGATTG
TTTCTCGCGGATGATCAGTTTGATGATTTCCGGCTTGTGATGACATTCAAGGATTCGA
AGATAACGGTCCAAGTGCCTTACCAGATTTCGACTTTCGGGATGTTGAAGATCTTCAGCT
AGCTGACTCTAGTTTCGGTTTTCTTGATCAACTTGCTCTATCAACATCTCTTGCCCAT
AAAAAGTTTTGCAGCTTCATAGGATCTTGCTTAGTAATGTTAAGTGAGAAGAGTGTTTTG
TTTTTTCGTTTATGCTTTAGTAATTAAGACATAAAAAGTGTGTGTTCCGGATTGTAGT
AAGATCTTAAGACATAAAGCCGGTTTTTGCAATTAGGAATCGAGTTTAAATGAAGTTTGA
GTTTATGTTTG

>G869 Amino Acid Sequence (domain in AA coordinates: 109-177)
MVAIRKEQSLSGVSEIKKRAKRNLTSSLPQETQPLRKVRIIVNDPYATDDSSSDEELK
VPKPRMKRIVREINFPMSMEVSEQPSSESSQDSTKTDGKIAVSASPAVPRKKPVGVRQRK
WGKWAABIRDPIKTRTWLGTFTLEEAAYDAKKLEFDAIVAGNVSTTKRDVSSSETS
QCSRSSPVVPEQDDTSASALTGVNPDVSTVAPTPTPNVPAGGNKETLFDLFDFTNLQ
IPDFGFLABEQDLDFDCFLADDQDFDLDDIQGFEDNGPSALPDFDFADVEDLQLAD
SSFGLDQLAPINISCLPSFAAS*

>G1645 (25..1104)

CGTCGACCTCCCAACACTAATCCATGTTTATAACGGAAAAACAAGTGTGGATGGATGAG
ATCGTCGCAAGAAGAGCTTCTTCTTCTTGGGACTTCCCTTTCAACGACATTAAATATTCAT
CAGCATCATCATCGTCACTGCAACACAAGTCATGAGTTTGAATCTTGAAGAGTCTCTT
GGAGATGTAGCGGTTACGAAGAAGAGAGTAATAATAAACCCTAATTTTCAGTAACAGC
GAGAGTGGTAAGAAGGAGACAACAGATAGTGGTCAGTCTTGGTCTCGTCGTCTTCAAAA
CCATCGGTCTTGGGGAGAGGACATTGGAGACCAGCTGAAGATGTTAACTCAAGAGACTT
GTCTCCATTACGGCCCAAAAACCTGGAACCTCATAGCTGAAAAGCTTCAAGGAAGATCT
GGGAAGAGCTGTAGACTACGATGGTTTAAACCAATTGGACCCGAGGATAAACCGAAGAGCT
TTCACAGAAGAAGAAGAGGAGAGGCTGATGCAAGCACATAGGCTTTATGGTAACAAATGG
GCAATGATTGCGAGGCTTTTCCCTGGTAGAAGTGAATTCAGTGAAGAACCATTGGCAT
GTTGTGATGGCTCCTAAGTATAGAGAACTCTTCTGCTTACCGTAGGAGAAAGCTTATG
AGTAATAATCCACTTAAACCTCACCTCACCAATAATCATATCCTAACCCTAACCCCTAAT
TACCACTCTTTTATCTCCACTAATCATTACTTCGCTCAGCCTTTCCCGAGTTTAAATTTG
ACTCATCACTGGTTTAAATAATGCCCCCTATCAGAGTGACCATAACCAGCTTGTGTGCTT
TTCCATTGCTTTCAAGGTTATGAGAACAATGAACCTCCGATGGTTGTGAGTATGTTTGGC

>G1038 (240..1574)

>G1038 Amino Acid Sequence (domain in AA coordinates: 198-247)

>G1073 (62..874)

CCCCCGACCTGCCTCTACAGAGACCTGAAGATTCCAGAACCCACCTGATCAAAAATAA
CATGGAACCTAACAGATCTGAAGCAGACGAAGCAAAGGCCGAGACCACTCCCACCGGTGG
AGCCACCAGCTCAGCCACAGCCTCTGGCTCTTCTCCGGACGTCGTCCACGTGGTCTGTC
TGCAGGTTCCAAAAACAAACCCAAACCTCCGACGATTATACTAGAGATAGTCTTAACGT
CCTTAGATCACACGTTCTTGAAGTCACCTCCGGTTCGGACATATCCGAGGCAGTCTCCAC
CTACGCCACTCGTCGGGCTGCGGCTTTGCAATTATAAGCGGCACGGGTGCGGTCACTAA
CGTCACGATACGGCAACCTGCGGCTCCGGCTGGTGGAGGTGTGATTACCGTCATGGTCCG
GTTTGACATTTTGTCTTTGACCGGTACTGCGCTTCCACCGCTGCACCACCGGGAGCAGG
AGGTTTGACGGTGTATCTAGCCGGAGGTCAAGGACAAGTTGTAGGAGGAATGTGGCTGG
TTCGTTAATTGCTTCCGGACCGGTAGTGTGATGGCTGCTTCTTTGCAAACGCAGTTTA
TGATAGGTTACCGATTGAAGAGGAAGAAACCCACCGCCGAGAACCACCGGGGTGCAGCA
GCAGCAGCCGGAGGCGTCTCAGTCGTCGGAGGTACGGGGAGTGGGGCCCAGGCGTGTGA
GTCAAACCTCCAAGTGGAAATGGTGGAGGAGGTGTGCTTTCTACAATCTTGAATGAA
TATGAACCAATTTTCAATTTCTCCGGGGGAGATATTACGGTATGAGCGGCGGTAGCGGAGG
AGGTGGTGGCGGTGCGACTAGACCCGCGTTTTAGAGTTTGTAGCGTTTGGTGACACCTTT
TGTTGCGTTTGCCTGTTTGACCTCAAACCTACTAGGCTACTAGCTATAGCGGTTGCGAAAT
GCGAATATTAGGTT

>G1073 Amino Acid Sequence (domain in AA coordinates: 33-42, 78-175)
MELNRSEADEAKAETPTGGATSSATASGSSSGRRPRGRPAGSKNPKPPTIITRDSNV
LRSHVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGR
FDILSLTGTALPPPAPPGAGGLTVYLAGGQGVVGGNVAGSLIASGPVVLMAASFANAVY
DRLPIEEEETPPPRITGVQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMN
MNNFQFSGGDIYMSGSGSGGGGGATRPAF*

>G1146 (129..3095)

cttctctagcgtcactcttcttcttctcattggctcggtagaataaggccaaggaagggatca
gttttaagttttgtttcattctttttagtgaggagaaaaagagttttgaaaatcaaaac
aacaataatgctgattaggcaaatgaaagatagctctgagactcacttagttatcaaaa
cccaacctttaagcaccacaatccaaaaaccgttcaaaacggtaaaatccctcctcctt
ctccttctcgggtgacggtgactactccggcgacggttactcagagtcaagcttcttcac
cttcaccaccgtcaaaagaatcgtagccggaggagaaaccgtggtggaagaaaatctgatc
aaggagatgtttgtatgagacctagctctcgtcctcgtaaaccgccaccgccaagtcaaa
ccacttctcgcgctctcgcgcgccaccgcccgtgagattgtcgctgtgaatcatcaga
tgcagatgggtgttcgtaaaaactcaaaactttgctccaagacctggatttggaacacttg
gaactaaatgcattgttaaagctaaaccactttctcgctgatttgacctaccaaggatttga
atcagtatgatgttacaattactcctgaagtgtcatcaaagagtggttaacagagctataa
ttgctgagtttagttagactttacaagagtgctgatctcgggaggagacttccggcttacg
atggccggaaaagtctttactgctggagaacttcttttacttggaaggagttcagtg
ttaagattgttgatgaagatgacggtatcatcaatggccctaaaaggagagatcatata
aggtggcaatcaagttgttgacggggcaaatatgcatcacttaggcgagtttctagctg
gtaaacgggcagattgtccgcaagaggcggtgcagattcttgatattgtactcagggagt
tgtcgggttaagaggttttgtcccggtggaagatcttcttcttcgcctgatattaaacac
cgcagcgactcgggtgaaggggttagagtcagtggtgtgggttttaccagagatttagaccaa
ctcaaatgggtttatcactaaatatcgatatggcttcagctgcattcatcgagcctcttc
cagtgatagagttttagcacagcttcttggaaggatgtcttgctgaagccattgtcgg
attctgatcgctcaagattaaagaagggctcttagaggagtgaagtagaggttactcaca
gagcgaatgtaagaaggaaataccgtgttgccgggtttaacaactcaaccaacaagagagc
taatgtttccagtagatgagaactgtactatgaagtcagttatttgagtatttccaagaga
tgtatggattcacgatccagcacacgcatttgccatgtctccaagttggaaccaaaga
aggcaagctatttgccgatggagcatgcaaaattgtcgagggaacacggtacacgaaaa
gggtgaatgagaagcagattactgctctcttgaaagttacatgccaaaggccgagggac
agagaaacgatatttgccgactgtccaacacacgcgatgatcaagatccatattgcaa
aggagtttgcatgaacataagcgaaggttagcttctgttgaaagctcgattcttccag
ctccatggcttaagtatcacgagaacgggaagaaaaagattgtctcccgcaagttggtc
agtgaatatgatgaacaagaaaatgatcaacgggatgactgtgagcagatgggcctgtg
ttaacttctcacgcagcgttcaagaaaacgttgctcggtgattttgtaatgaactgggtc
agatgtgtgaagctcagggcatggagtttaatccagaacccgtgataccaatatatagtg
cgaggcccgatcaagtcgagaaaagctctaaagcatgtttatcacacttcaatgaacaaaa

ccaaaggcaaagagtttagagcttctgctggcaatattacctgataacaacgggttcacttt
atggtgatcttaagagaatctgtgaaaccgagcttggtttgatattctcaatggtgtctca
caaacatgtgttcaagattagcaaacagtatctggcagatgtatcccttaaaatcaacg
taaagatgggaggaaggaacacagttcttagtagacgccataagctgtagaattccactgg
ttagcgataataccgacaatcatttttggcgagacgtgactcaccagagaacggggaag
agtcaagcccttcaatcgctgctgtgtgtgcttctcaagactggcctgaagtgaacaaat
atgcgggttagttgtgctcaagctcacaggcaagaacttatacaagatttgtataaaa
catggcaagatcctgttcgcggtactgttagtggcggtatgatcagggaaccttcttatct
catttagaaaagcaacagggaacaaaccgcttcgaattatctttatcgtgatggagtaa
gcgaagggaattctatcaagttttactctatgagttggatgcaattcgaaaggcttg
catcgcttgaaacgaattatcagccaccggtgacattcattagttgtacagaagcgccacc
acactcggttgttctgtaataatcaccgagacaaaaacagtactgaccgaagcggaaata
tcttaccaggtactgtagttgacactaaaaatgtcatccaactgaattcgacttctacc
ttttagccatgcgggtattcaggggaacagcaggcctgcacattaccatgttcttggg
acgagaacaatttcacagcagatggtattcaatctctgactaacaatctctgttatacct
atgcgcggtgactcggtcggtctctatagttcctccagcgtattatgctcatcttgag
catttcgagcacgtttctacctggaacctgagataatgcaagacaacggatcacgggta
aaaagaacacgaaacaacaactgtcggagacgtaggtgtgaagcctttaccagccttga
aggagaatgtgaagagagtaattgttctactgctaaaaatccaaacattccttaatcagtt
ttaataagtagtttgggtgttggctgttagttcggttttagatttaccatgttttctt
atgtaattttgtcggttgggttaagcctttaggaattagtgattagggtttttctaa
agttgtacttttagctgatgataacgttgatgcagtgactttgttaaacctcctcttcta
cagtagtggttacgtcggtcctc

>G1146 Amino Acid Sequence (domain in AA coordinates: 886-896)

MPIRQMKDSSETHLVIKTQPLKHHNPKTQVNGKIPPPSPSPVTVTTPATVTQSQASSPSP
PSKNRSRRNRGRKSDQGDVCMRPSSRPKPPPSQTTSSAVSVATAGEIVAVNHQMOM
GVRKNSNFAPRPGFGLTGTCKIVKANHFLADLPTKDLNQYDVTITPEVSSKSVNRALIAE
LVRLYKESDLGRRPLPAYDGRKSLYTAGELPFTWKEFSVKIVDEDDGIINGPKRERSYKVA
IKFVARANMHHLGEFLAGKRADCPQEAQVILDIVLRELSVKRFCPVGRSFFSPDIKTPQR
LGEGLSWCGFYQSIRPTQMGLSLNIDMASAAFIEPLPVIEFVAQLLGKDVLSKPLSDSD
RVKIKKGLRGVKEVTHRANVRKYRVAGLTTQPTRELMFPVDENCTMKSVIEYFQEMYG
FTIQHTHLPCLQVGNQKKASYLPMEACKIVEGQRYTKRLNEKQITALLKVTQRAEQRN
DILRTVQHNAYDQDPYAKEFGMNISEKLASVEARILPAPWLKYHENGKEKDCLPQVGQWN
MMNKKMINGMTVSRWACVNFSSVQENVARGFCNELGQMCEVSGMEFNPEPVIPIYSARP
DQVEKALKHVYHTSMNKTGKLELLAILPDNNGSLYGLDKRICETELGLISQCCILTKH
VFKISKQYLADVSLKINVKMGGRRNTVLVDAISCRIPLVSDIPTIIFGADVTHPENGEES
PSIAAVVASQDWPEVTKYAGLVCAQAHRELIDLYKTWQDPVRGTVSGGMIRDLLISFR
KATGQKPLRIIFYRDGVSEGFYQVLLYELDAIRKACASLEPNYQPPVTFIVVQKRHHTR
LFANNHRDKNSTDRGNILPGTVVDTKICHPTEFDLYLCSHAGIQGTSRPAHYHVLWDEN
NFTADGIQSLTNNLCYTYARCTRSVSIVPPAYYAHAAFRARFYLEPEIMQDNGSPGKKK
TKTTTVGDVGVKPLPALKENVKRMVFC*

>G1267 (152..967)

AAGTAGAGAATAATAATCACATCAAGATTGTTTATAACCCCTCCCNTAATCACCTTCTTA
NTNACCACCCTCTCCGGCTCTCAACAGAACAAACAAAAAACAGCTTCCGTTGTCCTG
TTCCGGCGAAATCGGACGGTCGAGATCAATCATGCATCGTAGAGCAGCAATTCAGAATC
GGATGACGAAGAAGATGAGACTTACAACGACGTCGTTCTGTAATCTCCTTCTTCTGTGA
AGACTCAAAGATCTCAAAACCAACTCCAAAGAAAAGGAGGAACGTAGAGAAGAGAGTTGT
CTCAGTTCCGATAGCTGACGTGGAAGGATCTAAGAGCAGAGGCGAAGTATATCCACCGTC
CGATTATGAGGCTGGAGAAAGTACGACAAAAACCGATCAAAGGCTCGCCTTATCCAG
GGGATATTACAGATGTAGTAGCTCAAAGGATGTCCGGCGAGGAAGCAGGTGGAGAGAAG
CCGTGTGGACCCCTTCTAAGCTTATGATTACTTACGCCTGCGACCACAATCACCTTTCCC
TTCTCTCCGCTAACACCAAATCCACACCGCTCCTCCGTCGTCCTCAAACCGCAAA
GAAAGAGGAAGAATACGAAGAGGAGGAAGAAGAACTAACCGTCACCGCCGAGAGGAACC
ACCGGCGGACTTGATCTAAGCCACGTAGACTCACCGTTGCTATTAGGCGGCTGCTACAG
CGAAATCGGAGAGTTCGGGTGGTTCTACGACGCTCGATCTCATCATCTGTTCTTC
GAATTTCTCGACGTAACCTTAGAGAGAGGTTTTTCAGTAGGCCAAGAGGAAGATGAGTC
TTTGTTCCGTGATCTCCGTGATTTACCTGATTGCGCCTCCGTGTTCCCGGTGGGACTGT

TGCGACGGAGGAGCAACATCGAAGATGTGATTTTGGCGCCATTCTTTCTGTGATAGTTC
TAGATGAGTTTGTGTGTGTAGCCAAAACCAAAGAAAAAACACAATTTTTTTATTTTCC
ACTGTAAAGGTGTATCAATGGTGGATTCAATTTTTTAAAAA

>G1267 Amino Acid Sequence (domain in AA coordinates: 70-127)
MHRRAAIQESDDEDETYNDVVPESPSSCEDSKISKPTPKRRNVKRVVSVPIADVEGS
KSRGEVYPPSDSWAWRKYGQKPIKGSYPYRGYYRCSSSKGCPARKQVERSRVDP SKLMIT
YACDHNHPFPSSSANTKSHRRSSVVLKTAKKEEYEEEEELTVTAEEPPAGLDLSDHVD
SPLLLGGCYSEIGEFWFYDASISSSSGSSNFLDVTLEGRFSGVQEEDESLFGDLGLDLPD
CASVFRRGTVATEEQHRRCDFGAIPFCDSSR*

>G1269 (88..951)
AACAATTCTCTCTCTTTATTCTTCTTCTCAGCTTCAGATTTTCAATCTTAAATCTTC
AAGTCTTCTTCTTCTTCTTCTGCAACCATGGCTATGCAGGAACGTTGTGAGAGTTTATGT
TCTGATGAACCTTATATCTTCTCAGATGCCCTTTTACCTCAAGACAAGAAAGCCTTATACC
ATCACTAAACAAAGAGAGAGAAATGGACAGAGCAGAGCATGAGAAGTTTGTAGAAGCATTG
AAACTCTATGGCAGAGCTTGGAGACGAATCGAAGAACATGTTGGAACAAAACTGCAGTT
CAGATTCGAAGCCATGCGCAGAGTTCTTTACTAAGGTTGCTCGCGATTTTGGTGTAGC
TCTGAGTCCATTGAGATCCCGCTCCAAGGCCAAAGAGAAAGCCGATGCATCCTTACCCT
AGAAAGCTTGTGATTCTGTATGCAAAAGAGATGGTATACGCTGAACCTAACCGGATCCAAG
CTGATTGAGATGAAGATAACCGATCTCCAACATCGGTTTTATCAGCTCATGGCTCAGAT
GGATTAGGTTCCATTGGTTCAAATTCACCTAACTCTTCTCAGCTGAGTTATCATCTCAC
ACAGAGGAATCAATTGTCTCTAGAAGCAGAGACCAACAGAGCCTTAAGCTCTTTGGAAAA
ACTTTTGTAGTTGGTGATTACAACCTTCAATGAGTTGTGATGATTCTGAAGATGGCAAG
AAGAAGCTATACTCAGAAACACAGTCTCTTCAATGTTCTTCTTCTACTTCAGAAAACGCT
GAAACAGAAAGTGGTAGTGTGCGAGTTCAAAGAAAGTGAGAGATCAGCTTTCTCTCAGTTA
AAATCGTGGTGACTGAGATGAACAACATGAGAGGGTTTATGCTTACAAAAAGAGAGTA
AAGGTGGAAGAAACATTGACAATGTAAATTTATCATATCCTTTGTGGTGAAGTGTTCGT
TTGTGTCAAGTCAGTTGTGTAACTCTTTTGATCTCAACATCAGATTATGTGTATAATGT
CAGATTATTAGGGAAAGTTTTTTTTGATTAGATTTCGTAAGATCACTCCAAAGTTTCGTGT
CTTTCCATATAACCAAGTTAGAAATTGAGATCCTTGTACTTAAACATTTTTATTGTATCAA
TCAAATCTTCTGATGAAAAA

>G1269 Amino Acid Sequence (domain in AA coordinates: 27-83)
MAMQERCELSDELISSSDAFYLKTRKPYTITKQREKWTEAEHEKFVEALKLYGRWRR
IEEHVGTKTAVQIRSHAQKFFTKVARDFGVSSSEIEIPPPRPKRKPMHPYPRKLVIIPDAK
EMVYAEITGSKLIQDEDNRSPTSVLSAHGSDGLSGISNSPNSSSSELSSHTESLSLEA
ETKQSLKLFKFTFVVGDNSSMSCDSEDKKKLYSETQSLQCSSSTSENAETEVVVSEF
KRSERSAFSQLKSSVTEMNMRGFMPYKRVKVEENIDNVKLSYPLW*

>G1452 (175..1296)
ATTTATTAAGCATCAATGAGAGAACTTCAGAGCTGGGTTTGTGTTCTGTCCAATAATACA
TAACCACGTTATCAATTTTTGTCTTTACTATCTCATTACACTCTTCTGTTATTCGCCCAA
TTCTTACAGTCATTACTCTCTATAGGGCTCGAGCGGCCCGCCGGCAGGTTTCTATGCAG
ATGGTTACACTTCCCGCTCCATTGCCAGATTGGGTTTGGTGTAAAGTCGCAATTAGTA
CTCACTATAGGGCTCGAGCGGCCCGCCGGCAGGTAAGATCAACAATGTCTAAAGAA
GCTGAGATGTCGATCGCGGTGTGCGCTTTGTTCCCTGGTTTTAGATTCTCTCTACTGAT
GTTGAACCTTATCTCGTACTATCTTCTGTCGTAATCGATGGTGTGAGAACTCTGTTGCT
GTGATTGCTGAGGTCGAGATTTACAAGTTTCGAGCCGTGGGACTTGCCAGAGGAATCGAAA
CTGAAATCGGAGAACGAGTGGTTTTACTTCTGCGCGAGGGGGAGGAAGTACCCGCACGGG
TCACAAAGCCGGCGAGCCACACAGCTAGGATATTGGAAGCGACCGGTAAGAGCGGAGT
GTTAAATCCGGGAACCAAGTTGTTGGAACCAAGAGAACGCTTGTATTTTATATCGGTGCG
GCTCCTCGTGGCGAGAGAACGGAGTGGATTATGCATGAATACTGCATCCATGGAGCCCCA
CAGGATGCATTAGTGGTGTGCGGTTAAGAAAAATGCTGATTTTGGGCTAGTTCGACC
CAAAAAATTGAGGATGGTGTGTGCAAGACGATGGCTACGTTGGCCAAAGAGGTGGTTTG
GACAAGGAGGACAAATCCTACTATGAATCTGAGCATCAGATACCAATGGTGACATCGCA
GAATCATCAAATGTTGTTGAGGATCAGGCCGATACCGATGATGATTGTTACGCCGAGATT
CTGAACGATGATATAATAAGCTCGACGAAGAAGCGTTGAAAGCTAGCCAAGCGTTTCGA
CCAATAATCCAACCTCATCAAGAAACAATATCAAGCGAGTCATCGAGTAAGAGGTCAAAA
TGTGGTATAAAAAAGAATCAACGGAAACAATGAATTGTTACGCTTTGTTTCAGGATCAAG
AACGTTGCCGGAACCGACTCCAGCTGGAGATTCCCGAACCCGTTCAAATCAAGAAAGAT

GATAGCCAGAGATTGATGAAGAATGTTCTGGCCACTACTGTTTCTTGGCTATCTTATTT
TCTTTCTTTTGGACTGTATTAATAGCTAGGAACCTAAAGCTAGTTACGACATACATATTAT
TTATACATAAATAAATATAGTATTTTGTCTATGGCAAAAAAAAAAAAAAAAAA

>G1452 Amino Acid Sequence (domain in AA coordinates: 30-177)
MQMVHTSRISIAQIGFVKSQLVLTIGLERPPGQVKDQMSKEAEMSIASVSLFPGFRFSP
TDVELISYLLRRKIDGDENSVAIVAEVEIYKFEPWDLPEESKLKSENEWFFYCARGRKYP
HGSQSRRTQLGYWKATGKERSVKSGNQVVGTKRTL VFHIGRAPRGERTEWIMHEYCIHG
APQDALVVCRLRNADFRASSTQKIEDGVVQDDGYVGQRGGLDKEDKSYESEHQIPNGD
IAESSNVVEDQADTDDCYAEILNDDIIKLDEEALKASQAFRPTNPTHQETISSESSSKR
SKCGIKKESTETMNCYALFRIKNVAGTDSWRFPNPFKIKKDDSQRLMKNVLATTVFLAI
LFSFFWTVLIARN*

>G1494 (114..1406)
TCGACAGAGTTGTGTTGGGCGTGGAACCTGGACTAGTTCACATATCAGGTTATATAGAT
CTTCTCTTTCAACTTCTGATTTCGTCCAGAAGCTTTCCTAATCTGAGATCTGACATGGAAC
ACCAAGGTTGGAGTTTGGAGGAGAATTATAGTTTGTCCACTAATAGAAGATCTATCAGGC
CACAAGATGAACCTAGTGGAGTTATTATGGCGAGATGGACAAGTGGTTCTGCAGAGCCAAA
CTCATAGAGAACAAACCCAAACCCAGAAACAAGATCATGAAGAAGCCCTAAGATCCA
GCACCTTTCTGAAGATCAAGAACTGTCTCTTGGATCCAATACCCCTCCAGATGAAGACC
CATTGCAACCCGACGACTTCTCTCCACTTCTTCTCAACCATGGATCCCTCCAGAGAC
CAACCTCAGAGACGGTTAAGCCTAAGTCCAGTCCCTGAACCTCCTCAAGTCATGGTTAAGC
CTAAGGCCCTGTCTGACCCCTCTCTCAAGTCATGCCTCTCCAAAATTTAGGTTAACAA
ATTATCATATCGGGGATTAGGGAACAGAAATGGAACAGTACTCGGTAACGACCGTTGGAC
CTAGCCATTGCGGAAGCAACCCATCACAGAACGATCTCGATGTCTCAATGAGTCATGATC
GAAGCAAAAACATAGAAGAAAAGCTTAATCCGAACGCAAGTTCTCATCAGGTGGCTCCT
CTGGTTGCAGCTTTGGCAAAGATATCAAAGAAATGGCTAGTGAAGATGCATCACAAACCG
ACCGTAAGAGAAAACGTATAAATCACACTGACGAATCTGTATCTCTATCAGATGCAATCG
GTAACAAGTCGAACCAACGATCAGGATCAAACCGAAGGAGTCGAGCAGCTGAAGTTTATA
ATCTCTCCGAAAGGAGGAGAGAGATAGGATCAATGAGAGAATGAAGGCTTTGCAAGAAC
TAATACCTCACTGCAGTAAACTGATAAAGCTTCGATTTTAGACGAAGCCATAGATTATT
TGAAATCACTTCAGTTACAGCTTCAAGTGATGTGGATGGGAGTGGAATGGCGGCGGCGG
CGGCTTCGGCTCCGATGATGTTCCCGGAGTTCAACCTCAGCAGTTTCATACGTGATAC
AGAGCCCGGTACAGTTACCTCGATTTCCGGTTATGGATCAGTCTGCAATTCAGAACAAATC
CCGGTTTAGTTTGCCAAAACCCGGTACAAAACCCAGATCATCTCCGACCGGTTTGCTAGAT
ACATCGGTGGGTTCCACACATGCAGGCCGCGACTCAGATGCAGCCGATGGAGATGTTGA
GATTTAGTTTACCGGCGGGACAGCAAAGTCAACAACCGTCGTCTGTGCCGACGAAGACCA
CCGACGGTTCTCGTTTGGACCACTAGGTTGGTGAGCCACTTTGC

>G1494 Amino Acid Sequence (domain in aa coordinates: 261-311)
MEHQGSFEENYSLSTNRRSIRPQDELVELLWRDQVVLQSQTTHREQTQTOKQDHHEEAL
RSSTFLEDQETVSWIQYPPDEDPFEPDDFSSSHFFSTMDPLQRPTSETVKPKSSPEPPQVM
VKPKACPDPPPQVMPPPKFRLTNSSSGIRETEMEQYSVTVGPSHCGSNPSQNDLDVSMS
HDRSKNIEEKLNPANSSSSGSSGCSFGKDIKEMASGRCTTDRKRKRINHTDESVSLS
AIGNKSNQRSGSNRRSRAAEVHNLSERRRRDRINERMKALQELIPHCSKTDKASILDEAI
DYLKSLQLQLQVMWMSGMAAAAAAPMMFPGVQPPQFIRQIQSPVQLPRFPVMDQSAIQ
NNPGLVCQNPVQNIISDRFARYIGGFPHMQAATQMQPMEMLRFS SPAGQQSQQPSSVPT
KTTDGSRLDH*

>G1548 (1..2511)
ATGGCAATGTCTTGCAAGGATGGTAAGTTGGGATGTTTGGATAATGGGAAGTATGTGAGG
TATACACCTGAACAAGTTGAAGCACTTGAGAGGCTTTATCATGACTGTCTAAACCGAGT
TCTATTCGCCGTGACAGTTGATCAGAGAGTGTCTATTCTCTTAACATTGAGCCTAAA
CAGATCAAAGTGTGTTTTCAGAACCGAAGATGTAGAGAGAAACAAAGGAAAGAGGCTTCA
CGGCTTCAAGCTGTGAATCGGAAGTTGACGGCAATGAACAAGCTCTTGATGGAGGAGAAT
GACAGTTGACAGAAGCAAGTGTACAGCTGGTCCATGAAAACAGCTACTTCCGTCAACAT
ACTCCAAATCCTTCACTCCAGCTAAAGACACAAGCTGTGAATCGGTGGTGACGAGTGGT
CAGCACCAATTGGCATCTCAAAATCCTCAGAGAGATGCTAGTCTGCAGGACTTTTGTCC
ATTGCAGAAGAACTTTAGCAGAGTTTCTTTCAAAGGCAACTGGAACCGCTGTGAGTGG
GTTTCAGATGCCTGGATGAAGCCTGGTCCGATTCATTGAATCATCGCTATTTCTCAT
GGTTGCACTGGTGTGGCAGCAGCGCCTGTGGCCTAGTGGGTCTTGAGCCTACAAGGGTT

GCAGAGATTGTCAAGGATCGTCCTTCGTGGTTCCGCGAATGTCGAGCTGTTGAAGTTATG
AACGTGTTGCCAACTGCCAATGGTGGAACCGTTGAGCTGCTTTATATGCAGCTCTATGCA
CCAACCTACATTGGCCCCACCACGCGATTCTCGGCTGTTACGTTACACCTCTGTTTTAGAA
GATGGCAGCCTTGTGGTGTGCGAGAGATCTCTTAAGAGCACTCAAAATGGTCCTAGTATG
CCACTGGTTTCAGAAATTTGTGAGAGCAGAGATGCTTTCCAGTGGGTACTTGATACGGCCT
TGTGATGGTGGTGGCTCAATCATACATAGTGGATCATATGGATTTGGAGGCTTGTAGC
GTGCCTGAGGTCCTTGCGCCCGCTCTATGAGTCACCCAAAGTACTTGACAGAGACAACA
ATGGCGGCACATGCGTCAGCTCAAGCAAATAGCTCAGGAGGTTACTCAGACTAATAGTAGT
GTTAATGGGTGGGACGGCGCTCTGCTGCCCTTAAGAGCTCTCAGCCAGAGGCTAAGCAGA
GGCTTCAATGAAGCTGTAAATGGTTTCACTGATGAAGGATGCTCAGTGATAGGAGATAGC
ATGGATGATGTACAATCACTGTAAACTCTTCTCCAGACAAGCTAATGGGTCTAAATCTT
ACATTTGCCAATGGCTTTGCTCCTGTAAAGCAATGTTGTTTTATGCGCAAAAGCATCAATG
CTTTTACAGAATGTTCTCCGCGCATCTGCTTCGGTTTCTGAGGGAGCATAGGTCAGAA
TGGGCTGACAACAACATTTGATGCGTATCTAGCAGCAGCAGTTAAAGTAGGGCCTTGTAGT
GCCCCGAGTTGGAGGATTTGGAGGGCAGGTTATACTTCCACTTGCTCATACTATTGAGCAT
GAAGAGTTTATGGAAGTCATCAAATTGGAAGGTCTTGGTCATTCCCCTGAAGATGCAATC
GTTCCAAGAGATATCTTCTTCTCAACTTTGTAGCGGAATGGATGAAAATGCTGTAGGA
ACCTGTGCGGAACCTATATTTGCTCCAATCGATGCTTCGTTTGC GGATGATGCACCTCTG
CTTCTTCTGGTTTTCTGATTTATCCCTCTTGATTCCGCAAAAGGAAGTATCTAGCCCCAAC
CGAACCTTGGATCTGCTTCGGCACTGGAAATGGTTTCACTGGAACAAAAGCCTCAACT
GATCAATCAGGAACTCCACATGTGCAAGATCTGTGATGACAATAGCATTGAGTTTGGT
ATCGAGAGCCATATGCAAGAACATGTAGCATCCATGGCTAGGCAGTATGTTGAGGTATC
ATATCATCGGTGCAGAGAGTAGCATTGGCTCTTCTCTCTCATATCAGCTCACAAGTT
GGTCTACGCACTCTTTGGGTACTCCTGAAGCCCAAACACTTGCTCGTTGGATTTGCCAG
AGTTACAGGGGCTACATGGGTGTTGAGCTACTTAAATCAAACAGTGACGGAATGAATCT
ATTCTTAAGAATCTTTGGCATCACACTGATGCTATAATCTGCTGCTCAATGAAGGCCTTG
CCCGTCTTACATTGCAAACCAGGCGGACTTGACATGCTGGAGACTACATTAGTTGCT
CTTCAAGACATCTCTTTAGAGAAGATATTTGATGACAATGGAAGAAAGACTCTTTGCTCT
GAGTTCCACAGATCATGCAACAGGGCTTCGCGTGCCTTCAAGGCGGGATATGTCTCTCA
AGCATGGGGAGACCAGTTTCGTATGAGAGAGCAGTTGCTTGGAAAGTACTCAATGAAGAA
GAAAATGCTCATTCATCTGCTTTGTGTTTCAATGGTCTTTGTGTGA

>G1548 Amino Acid Sequence (domain in AA coordinates: 17-77)
MAMSKDGLKGLDNGKYVRYTPEQVEALERLYHDCPKPSSIRRQQLIRECPILSNIEPK
QIKVWFQNRRCREKQKREASRLQAVNRKLTAMNKLMEENDRLQKQVSQLVHENSYFRQH
TPNPSLPARRKDTSCSVVTSQHQQLASQNPQRDASPAGLLSIAEETLAEFLSKATGTAVEW
VQMPGMPKPGPDSIGIIAISHGCTGVAARACGLVGLPEPTRVAEIVKDRPSWFRECRAVEVM
NVLPTANGGTVELLYMQLYAPTTLAPPRDFWLLRYTSVLEDGSLVVCERSLKSTQNGPSM
PLVQNFVRAEMLSSGYLIRPCDGGGSIIHIVDHMDLEACSVPEVLRPLYESPKVLAQKTT
MAALRQLKQIAQEVTQTNSSVNGWRRPAAALRALSQLSRGFNEAVNGFTDEGWSVIGDS
MDDVTITVNSSPDKMLGLNLTFFANGFAPVSNVVLCAKASMLLQNVPPAILLRFLREHRSE
WADNNIDAYLAAAVKVGPCSRVGGFGGVILPLAHTIEHEEFMEVIKLEGLGHSPEDAI
VPRDIFLLQLCSGMDENAVGTCAELIFAPIDASFADDAPLLPSGFRIIPLDSAKEVSSPN
RTLDLASALEIGSAGTKASTDQSGNSTCARSVMTIAFEFGIESHMQEHVASMARQYVRGI
ISSVQRVALALSPSHISSQVGLRTPLTPEAQTARWICQSYRGYMGVELLKSNSDGNES
ILKNLWHHTDAIICCSMKALPVFTFANQAGLDMLETTLVALQDISLEKIFDDNGRKTLC
EFPQIMQOGFACLOGGICLSSMRPVSYERAVAWKVLNEENAHICFVFINWSFV*

>G1574 (1..1962)

ATGGATGATACAATGGACATGAGTTCAGGTAGTGATGAAGAAGTACAAGAAGAGAAGACC
ACTGTTAACGAGAGGGTCACTATCAGGCTGCATTACAAGATCTGAAGCAACCCAAGACC
GAAAAGGATCTACCTCCTGGTGTCTTACAGTTCTCTTATGAGGCATCAGAAAATTGCA
TTGAACTGGATGCGTAAGAAAGAAAAAGAACAGGCAGGCACTGTTGGGAGGGATATTAGCA
GATGATCAGGGACTTGGTAAACGATCTCGACGATCTCTCTTATCCTGTTACAAAAGTTG
AAGTCACAATCAAAGCAGAGAAAGCGAAAAGGTCAAAACTCTGGTGGTACATTGATTGTT
TGTCACAGCAAGTGTGTAAACAATGGGCAAGAGAAGTTAAAGAGAAGGTTTCTGATGAA
CACAAACTCTCTGTTTGTAGTCCACCATGGATCTCACAGAACCAAGATCCAACAGAAATA
GCAATATATGATGTGGTCATGACAACCTTACGCCATTGTTACAAATGAAGTTCCACAAAAC
CCTATGCTGAATCGTTATGATAGTATGAGAGGCAGAGAAAGCCTTGACGGATCGAGTTTG

ATTACAGCTCAGCTTGGTGCCTAGGAAGAGTTAGGTGGTTGAGAGTAGTATTAGATGAA
 GCTCATACAATTAAAAACCATAGAACCTAATTGCAAAAGCTTGTCTTAGCCTTAGAGCC
 AAAAGGAGATGGTGTGACTGGAACGCCGATAAAGAACAAGTAGACGATCTTTATAGC
 TATTTAGATTTCTTAGATATCATCCATATGCCATGTGCAATTCAATTCACCAAAGAATC
 AAAGCTCCAATTGATAAAAGCCTCTTCATGGTTACAAGAAGCTTCAAGCTATTCTAAGG
 GGTATAATGTTGCGCCGACCAAGAATGGTCTTTCTACAGGAAGCTTGAATTGAATTCA
 CGTTGGAAGTTTGAGGAATATGCTGCTGATGGGACTTTGCATGAACACATGGCTTATCTT
 TTGGTGATGCTTTTTCGACTACGCCAAGCTTGTAAACCATCCACAACCTTGTAAACGGATAT
 AGTCACTCAGATACTACAAGAAAAATGTCAGATGGAGTTCGAGTAGCCCCCTAGAGAGAAT
 CTAATCATGTTCCCTCGATCTCTTGAAATTATCCTCAACCACCTGCTCTGTTTGTAGTGAT
 CCACCAAAGACCTGTTGTTACTTTGTGTGGCCATGTGTTTTGTATGAGTGTGTGTCT
 GTAAACATTAAACGGGGATAACAATACGTGCCCTGCACCTTAATTGCCACAGCCAGCTTAAA
 CATGATGTTGTTTTCACTGAATCTGCAGTTAGAAGTTGCATCAACGATTATGATGATCCT
 GAAGATAAAATGCTTTAGTTGCATCAAGGCGAGTTTATTTTCATCGAAAATCCGAGCTGT
 GATAGAGATTTCTTCAGCTGCTTGCAGAGCAAGGCAGTCCAGACACTCCACCAATAAGAGC
 AATAGTATCAGTGGACTGAATCTCATTTTTACGTTTCTCAAAGACAAATGTAATGATTAT
 GAAACAGGTGCGATGTTGATGTCTCTTAAAGCTGGAAACCTTGGATTGAATATGGTAGCT
 GCAAGTCATGTCTTCTACTGGACCTATGGTGAATCCAACAACAGAGGATCAAGCTATT
 GATCGAGCTCATCGTATCGGACAAACTCGAGCTGTACGGTCACCTCGTATTGCCATCAAA
 AATACCGTTGAGGAACGAATTTTGACTCTTCATGAACGTAAAAGGAACATTGTTGCATCT
 GCATTGGGTGAAAAAACTGGCAAAAGTTCTGCGATTCAACTAACACTAGAAGATCTCGA
 ATATCTGTTTTTTGGTGTGTAGAATATCCAGAGTTTTTTATTGATAAGAGGAATAAAACC
 TTTAGCTATTTAATAAGTCACAAGTGTGAATGTAATGAATAA

>G1574 Amino Acid Sequence (domain in AA coordinates: 28-350)
 MDDTMDSSGSDEEVQBEKTTVNERVIYQALQDLKQPKTEKDLPPGVLTVPMLMRHQKIA
 LNWMRKKEKRSRHLGILADDQGLGKTIISTISLILLQKLKSQSQQRKRKGQNSGGTLIV
 CPASVVKQWAREVKEKVSDEHKLVLVHGHSHRTKDPTEIAIYDVVMTTYAIIVTNEVPQN
 PMLNRYDSMRGRESLDGSSLIQPHVGALGRVRLRVVLDEAHTIKNHRTLIAKACFSLR
 KRRWCLTGTPIKNKVDDLYSYFRFLRYHPYAMCNSFHQRIKAPIDKKPLHGYYKKLQAILR
 GIMLRRTKEWSFYRKLELNSRWKFEEYAADGTLHEHMAYLLVMLLRRLRQACNHPQLVNGY
 SHSDTTRKMSDGVVRVAPRENLMFLDLLKLSSTTCSVCSDDPPKDPVVTLCGHVFCYECVS
 VNINGDNNTCPALNCHSQLKHDVVFTESAVRSCINDYDDPEDKNALVASRRVYFIENPSC
 DRDSSVACRARQSRHSTNKNDSISGLNLI FTFLKDKCNDYETGAMLMSLKAGNLGLNMVA
 ASHVILLDLWNPTTEDQAIDRAHRIGQTRAVTVTRIAIKNTVEERILTLHERKRNIVAS
 ALGEKNWQKFCDSNTRRSRISVFWCVEYPRVFDIKRNKTFSYLISHKCECNE*

>G1586 (1..807)

ATGAATCAAGAAGGTGCTTCACATAGCCCATCCTCCACTTCCACCGAACCAGTCCGGGCA
 CGTTGGTCACTAAACCGGAGCAAATCTTGATACTCGAATCCATCTTCAACAGTGGTACT
 GTTAACCCACCAAAAGATGAAACGGTGAGGATAAGAAAGATGCTTGAGAAATTCGGTGCT
 GTGGGAGACGCAACGCTCTTCTACTGGTTTCAAACCGACGGTCAAGATCTCGCCGGAGA
 CACCGGCAGCTTTTAGCAGCCACCACCGCAGCCGCCACCTCCATAGGAGCTGAAGACCAC
 CAGCACATGACGGCCATGAGCATGCATCAATATCCTTGCAGCAACAACGAGATTGATTTG
 GGGTTTGGAAGTTGTAGCAACTTATCAGCTAATTACTTCTTAAATGGATCGTCGTCATCT
 CAAATCCCTTCTTTTCTCCTCGGCCTCTCTTCTTCAAGTGGTGGGTGTGAGAACAAAT
 GGTATGGAGAATCTCTTCAAATGTATGGCCATGAATCTGATCATAATCATCAGCAGCAG
 CATCATAGCTCAAATGCTGCATCAGTTTAAACCCATCTGATCAAACTCCAACCTCCAA
 TACGAACAAGAAGGGTTTATGACGGTGTATATAACGGAGTTCCTATGGAAGTAACAAA
 GGAGCAATAGACATGAAAACAATGTTCCGGTGATGATTCCGGTGTACTTCATTCTCTGGT
 CTTCCTCTTCCCACTGATGAGTTTGGTTTCTTGTATGCATTCTTTACAACATGGACAAAT
 TATTTCTGTTACCGAGACAGACATGA

>G1586 Amino Acid Sequence (domain in AA coordinates: 21-81)
 MNQEGASHSPSSTSTEPVRARWSPKPEQILILESIFNSGTVPNPKDETVRIRKMLEKFGA
 VGDANVFYWFQNRRSRRRRHQLAATTAAATSIGAEDHQHMTAMSMHQYPCSNNEIDL
 GFGSCSNLSANYFLNGSSSSQIPSFLLGLSSSSGCGENNGMENLFKMYGHESDHNHQQQ
 HHSSNAASVLNPSDQNSNSQYEQEGFMTVFINGVPEVTKGAIDMKTMFGDDSVLLHSSG
 LPLPTDEFGFLMHSLOHGQTYFLVPRQT*

>G1786 (1..1170)

ATGATCGTGTACGGTGGGGGAGCATCCGAGGACGGTGAAGGTGGAGGGGTGGTTCTGAAG
AAAGGGCCATGGACGGTGGCCGAGGACGAGACACTGGCGGCTTACGTACGGGAATACGGT
GAAGGGAACTGGAATTCTGTTTCAAGAAGACATGGCTGGCTAGGTGTGGCAAGAGCTGC
CGCCTCCGCTGGGCTAACCACTTACGACCTAATCTCAGGAAAGGCTCCTTACCCCCGAG
GAAGAAGCTCTCATATACAACCTCCACTCTCAGCTAGGCAACAAATGGGCTCGCATGGCT
GCTCAGTTACCAGGCAGAACAGATAACGAGATCAAGAACTACTGGAACACGAGGTTGAAA
CGCTTCCAACGCCAAGGCCTCCCTCTCTACCTCCAGAATATTTCCAAAACAATCATCAA
CAACAAATGTATCCTCAACAGCCCTCCTCACCTCTCCGTCCTCAACACCTGCTTCTTCC
TTTACCTTTCTCTCTCCCAACCGCTTCTCTGTGTCCCAACGTTGTTATAACACTGCC
TTCTCTCCCAAGGCCTCATATATTTCTCTCTCAACCAATTTCTTGTCTCGTCTCCGACC
TTTCTTACACCCATTCTCTCTTTCTCTCTATCAGTCTACCAATCCGGTTTACTCCATG
AAACATGAGTCTCTTCAACCAAAATTCATACTCTGCCTCTTTAGGAGTCTATCAAGTA
AGCAAGTTCTCAGACAATGGGGATTGTAACCAAAACCTGAACACCGGTTTGCATACAAAT
ACCTGTCAGCTGTTAGAGGATCTTATGGAGGAGGCCGAGGCTCTAGCTGATAGCTTTCTGT
GCTCCTAAGCGGAGACAAATCATGGCTGCGCTTGAGGACAACAACAACAACAACATTTT
TTCTCGGGAGGTTTTCGGACATCGTGTCTTCTTCAACAGTCTATGTTCTTGCAGGTTTA
ACACCAAAGGAAGATGAGTCTCTCCAGATGAACACAATGCAAGATGAGGACATAACAAAG
CTTCTTGACTGGGGAAGTGAAGTGAAGAAATCTCAACGGGCAATCCTCTGTGATAACA
ACAGAGAACAACCTTGTCTTGCAGATCACCAGTTCGCTTTTCTGTTTCCAGTTGATGAT
GACACCAACAACCTTGCCAGGGATCTGCTAG

>G1786 Amino Acid Sequence (domain in AA coordinates: TBD)
MIVYGGGASEDGEVGGVVLKGPWTVAEDELAAAYVREYGEENWNSVQKKTWLRACGKSC
RLRWANHLRLPNLRKGSFTPEERLIIQLHSQLGKWARMAAQLPGRITDNEIKNYWNTRLK
RFQRQGLPLYPPEYSQNNHQQMYPQPPSSPLPSQTPASSFTFPLLPQPSLCPKRCYNTA
FSPKASYISSPTNFLVSSPTFLHTHSSLSYQSTNPVYSMKHELSSNQIPYSASLGVYQV
SKFSDNGDCNQNLNTGLHTNTCQLLEDLMEEAEALADSFRAKRRQIMAALEDNNNNNNF
FSGGFGHRVSSNSLCSLQGLTPKEDESLOMNTMQDEDITKLLDWGSESEEBISNGQSSVIT
TENNLVLDDHQFAFLFPVDDDTNNLPGIC*

>G1792 (77..496)

AATCCATAGATCTCTTATTAATAACAGTGCTGACCAAGCTCTTACAAAGCAAACCAATC
TAGAACACCAAAGTTAATGGAGAGCTCAAACAGGAGCAGCAACAACCAATCACAAGATGA
CAAGCAAGCTCGTTTCCGGGGAGTTTCAAGAAGGCCTTGGGGAAAGTTTGCAGCAGAGAT
TCGAGACCCGTCGAGAAACGGTGCCCGTCTTTGGCTCGGGACATTTGAGACCGCTGAGGA
GGCAGCAAGGGCTTATGACCGAGCAGCCTTTAACCTTAGGGGTCTATCTCGCTATACTCAA
CTTCCCTAATGAGTATTATCCACGTATGAGCAGTACTCTCGCTTCGCCCTCCTTATGCTTC
TTCTTCTTCTCGTCTCGTCAATCGGGTTCACTTCTACTAATGTGAGTCGACAAAACCAAG
AGAAGTTTTCGAGTTTGTAGTATTTGGACGATAAGGTTCTTGAAGAACTTCTTGATTGAGA
AGAAAGGAAGAGATAATCACGATTAGTTTTGTTTTGATATTTTATGTGGCACTGTTGTGG
CTACCTACGTGATTATGTGCATGTATAGGTGCTGATTAGTACTTTATAACATGCATG
CCACGACCATAAATTGTAAGAGAAGACGTACTTTGCGTTTTTCATGAAATATGAATGTTAG
ATGGTTTGAGTACAAAAA

>G1792 Amino Acid Sequence (domain in aa coordinates: 17-85)
MESSNRSSNNQSQDDKQARFRGVRRRPWGKFAAEIRDPSRNGARLWLGTFFETAEEAARAY
DRAAFNLRGHLAILNFPNEYPRMDYSLRPPYASSSSSSSGSTSTNVSRQNRQEVFEF
EYLDDKVLEBLDSEERKR*

>G1865 (48..899)

AAGAAGAGGACATGAAGCACAGAGATTCTGCAGACTGCAGGTGACCAATGGACACTTTAT
CAATAAAACATACCTACTACTCTCTTACACTTTCAATTTTCCAATACAAATCCCAATCT
TTAATCTCTCTTTCTTCTCATCTCTTCTCTTCTCTTCTCTTCTCATGGCTACAAGGATTC
CAATTCACAGAAATCAATGGGAAGAACTTGAAGAACCAAGCTCTTGTGTTCAAGTACTTAG
CTGCAATATGCCTGTTCCACCTCATCTTCTCTTCTCATCAAAAGACCCCTTCTCTTCT
CTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
TTGGGTGGAATGTGTATGAGATGGGAATGGGAAGAAAGATAGATGCAGAGCCAGGAAGAT
GTAGAAGAACTGATGGCAAGAAATGGAGATGCTCTAAGAAAGCTTACCCTGACTCTAAGT
ACTGTGAGAGACATGCATAGAGGCAAGAACCGTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
CTCAATTCACCTCAAAATCTCTTCTCGACTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
TGGATGATTTCTTCTCATAGAACCTTCCGGGTCAATCAAAGCTGCTCTGGCTCAGCAA

TGGAAGATAATGATGATGGCTCATGTAGAGGCATCAACAACGAGGAGAAGCAGCCGGATC
GACATTGCTTCATCCTTGGTACTGACTTGAGGACACGTGAGAGGCCATTGATGTTAGAGG
AGAAGCTGAAACAAAGAGATCATGATAATGAAGAAGAGCAAGGAAGCAAGAGGTTTATA
GGTTTCTTGATGAATGGCCTTCTTCTAAATCTTCTGTTTCTACTTCACTCTTCATTTGAT
CATCTTTTGTCTTATAACCTTGTATTTCTTGTTAAGATGGTAATGCAAATT
>G1865 Amino Acid Sequence (domain in AA coordinates: 124-149)
MDTLSIKTYLLLSYTFNFIQIPFNLSFFFISLSLSLFMATRIPFTESQWEELENQALV
FKYLAANMPVPPHLLFLIKRPFLFSSSSSSSSSSSFFSPTLSPHFGWNVYEMGMGRKIDA
EPGRCRRTDGKKWRCSEAYPDSKYCERHMRGKNRSSSRKPPPTQFTPNFLDSSSRRR
RSGYMDFFSIEPSGSIKSCSGSAMEDNDDGSCRGINNEEKQPDHCFILGTDLRLTRERP
LMLEEKLKQRDHDNEEQGSKRFYRFLDEWPSSKSSVSTSLFI*
>G1886 (43..909)
AGGAAACATAAGTAATCGTTGCTTCGATCCTTTGTACATGGATGGATCCTGAACAGGAA
ATCTCAAACGAGACTTTGGAACTATATTGGTAAGTTCAACAAAAGGAAGCAATAATAAC
AATAAGAAAATGGAAGAAGAAATGAAGAAGAAAGTATCAAGAGGAGAATTAGGAGGTGAA
GCTCAAAATTGTCCAAGATGTGAATCTCCAAACACAAAGTTTTGTACTACAACAACATAT
AGTCTCTCACAACTCGTTACTTCTGCAAATCTTGTGCGAGATATGGACTAAAGGCGGT
ACTCTTCGTAACGTTCCCGTCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT
GCTTTCTCCAAGAACAACAATAAGTCTATTAATTTCCATACTGATCCACTTCAGAAC
CCTTTAATTACGGGAATGCCACCATCATCTTTTGGTTATGATCACTCCATTGATCTCAAC
CTCGCTTCGCTACTCTCCAAAAGCATCATTTATCCTCTCAAGCTACTACGCCTTCTTTT
GGGTTTGGAGGTGATCTTTCTATTTATGGAACCTCAACGAATGATGTAGGGATCTTCGGA
GGGCAAAACGGTACTTATAACAATAGTTTGTGTTATGGGTTTATGTCCGGAATGGTAAT
AATAATCAAAATGAAATCAAGATGGCTTCTACATTGGGGATGTCTTTGGAAGGAACGAG
AGAAAGCAAGAGAATGTGAACAATAACAATAAATACTCAGAGAATCCTAGCAAGGTGTTT
TGGGGGTTTCCATGGCAGATGACCGGAGATTCCGCCGGAGTTGTACCGGAGATTGATCCC
GGAAGGGAAAGCTGGAATGGGATGCTTTCATCTTGAATAATGGTTTACTCAACACTCCT
TTGGTCTAGCAGATCATTAA
>G1886 Amino Acid Sequence (domain in aa coordinates: 17-59)
MDPEQEISNETLETILVSSTKGSNNNNKMEEMKKKVSRLGELGGEAQNCPRCESPNTKF
CYNNYSLSQPRYFCKSCRRYWKGGTLRNVPVGGGCRNRKRSSSAFSSKNNNNKSINFH
TDPLQNPLITGMPPSSFGYDHSIDLNLAFATLQKHHLLSSQATTPSFGFGDLSIYGNSTN
DVGIFGGQNGTYNNSLCYGFMSGNGNNNNQNEIKMASTLGMSLEGNERKQENVNNNNNNSE
NPSKVFWGFPWQMTGDSAGVVEIDPGRESWNGMVSSWNNGLNLTPLV*
>G1933 (33..1418)
AATTGAGATTAAAGTAATTTATCTTTCAGAAAATGGCGGTGAAGACGATGTATCTTTGA
TAAGAACGAGACGTTAGTGGCACCACAAAGACCCACGATTACAGTTCCTCATAGACCTC
CGGCGATCGAAGACGGCGGCTATTTCTTGGCGGTGGAGATGGGCTTAGTCTAAGCCAG
GGCCACTTTCTTTTGTCTCTTCTTTGTTTGTGATAACTTCCCTGACGTCTTGACGCCGG
ATAACCAACGACGACGTCGTTTACTCAGCTTCTTAACGGAATATGTCCGTGTCTCCTG
GTGGCGGAGGACGTTCAACGGCGGGGATGTTCCGCCGAGGAGTCCGATGTTTACAATCC
CTTCTGGTTTCAGCCCTTCTAGTCTTCTCACCTCGCCCATGTTCTTTCCCCCGCAGTCGT
CAGCTCATACCGGCTTTATTCAACCACGGCAGCAGTCACAACCGCAACCACAACGACCAG
ACACGTTTCTCACCATATGCCACCATCGACATCCGTCGCCGTCCATGGTCGTCAATCTT
TAGACGTTTCAACAAGTAGATCAAAGAGCTCGAAACCATATAATAATCCGGGGAATAACA
ATAATAACCGGTGCTATAACGTTGTGAACGTTGATAAACCGGCGGATGACGGTTATAACT
GGAGGAAGTACGGACAAAAGCCTATCAAAGGGTGTGAATATCCAAGGAGTTATTACAAAT
GTACACATGTTAACTGTCCGGTGAAGAAGAAAGTGAACGGTCATCGGATGGACAGATCA
CTCAGATCATTTACAAAGGTCAACATGATCACGAGAGGCCTCAGAATCGCCGTGGCGGTG
GAGGCAGAGATTCCACTGAGGTTGGTGGTGCAGGGCAAATGATGGAATCTAGTGATGATA
GTGGTTATCGTAAGGATCATGATGATGATGATGATGATGATGAAGATGATGAAGATCTTC
CGGCTTCAAGATAAGAAGATAGACGGTGTGTGTCGACGACTCACGGACGGTGACCGAGC
CTAAGATTATCGTTTCAGACAAAAGTGAAGTCAATCTTCTCGACGATGGCTATAGGTGGC
GTAAGTACGGACAAAAGTGTCAAAGGAAATCCCCATCCAAGGAGCTATTATAAATGTA
CAACGCCAAATTGTACGGTCCGTAAACATGTAGAGAGAGCTTCCACGGATGCTAAGGCTG
TGATTACAACCTTACGAAGGTAAACACAATCACGATGTCCCTGCCGCTAGAAACGGTACCG
CGGCAGCAACCGCAGCTCGGGTGGGGCCGTCTGACCACCATCGTATGAGATCAATGTCCG

GGAACAATATGCAACAACATATGAGTTTCGGTAACAATAAACACAGGCCAATCTCCGG
TTCTTTTGAGGTTGAAAGAAGAGAAAATCACAATTTGACTTTTAAGAACCAAGATTTTCG
AGATTGATATT

>G1933 Amino Acid Sequence (conserved domain in AA coordinates:205-263, 344-404)

MAVEDDVSLIRTITVAPTRPTITVPHRPPAIETAYFFGGGDLGLSLSPGPLSFVSSSLFV
DNFPDVLTPDNQRTTSFTQLNGTMSVSPGGGGRSTAGMFAGGGPMFTIPSGFSPSSLLT
SPMFFPPQSSAHTGFIQPRQSQPQRPDTPPHMPPSTSVAVHGRQSLDVSQVDQRRAR
NHYNNPGNNNNNRSYNNVNDKPADDGYNWRKYGQKPIKGCEYPRSYKCTHVNCVPVKKK
VERSSDGQITQIIYKGQHDHERPQNRGGGGRDSTEVGAGQMMESSDDSGYRKDHDDDD
DDDEDDDELPAKIRRIDGVSTTHRTVTEPKIIIVQTKSEVDLLDDGYRWRKYGQKVVKGN
PHPRSYKCTTPNCTVRKHVERASTDAKAVITTYEGKHNHDVPAARNGTAAATAAAVGPS
DHHRMRSMGNNMQHMSFGNNNNTGQSPVLLRLKEEKITI*

>G2059 (58..1089)

TTAAGAACAGGCTTCATCTCTGACAAACACTCAAAAAACAAACAAAAAGGAACATG
GAAGATCAGTTTCTTAAATAGAACTAGCTTCATGCACGACAAGCTCTTGCTCTCTGGA
ATCTACGGGTTCTTGAGTTCTTCGACGCGCCACAACCTCTCGGTGTTCGAATATTTTGG
GAAGGTATGAAATCTCTCTCTCTCTGCTTCTTCGACTCCGAGCTACTTTGTGTCGCCT
CATGATCATGAGCTCACATCTTCTATTCATCCATCTCCGGTAGCTTCTGTTCCTTGAAC
TTTCTAGAATCTTTCTCTCAGTCTCAACATCTCTGATCATCTCTTCTAAACCTCCAAAC
CTTACTTTGTTCCTTAAAGAACCAAGCTACTAGAAGCTTTCTCAATCCGAAAGCAACATG
AGCCCTTACCATAAAATACATCCAACTCCTTTTATCAATCAGACCAAAACAGAAACGAA
TGGGTAGAGATCAATAAACTCTAACCAACTATCCCTCGAAAGGTTTGGAAACTATTGG
CTAAGTACCACCAAGACTCAACCCATGAAGTCAAAAAACAAGAAAGGTTGTTGAGACGACG
ACCCCAACAAACTGTATAGAGGAGTGAGACAAAGACACTGGGGCAAATGGGTGCGAGAG
ATTAGGCTTCCAAGGAACAGAACCCGTGTTTGGCTCGGCACCTTTGAAACCGCTGAGCAA
GCAGCAATGGCTTACGATACAGCAGCTTATATCTTCTGTCGGAATTTCGCACACCTCAAC
TTTCTGATCTTAAACACCAGCTCAAGTCCGGTTCTTTGCGATGCATGATCGCCTCACTT
TTGGAGTCCAAGATTCAACAGATCTCATCTTCCCAAGTAAGTAAGTCTCTCTCTCTCTCT
CCTCCAAAGTGGGAACACCGGAGCAAAAGAATCATCATGAGATGGAGTCAGGAGAA
GACGTGATGATGAAGAAACAGAAAGCCATAAGGAAGTGAAGGAGATGGTGTACAA
TTGAGTAGGATGCCTTCTTTGGATATGGATCTCATTTGGGATGCTCTCTCATTTCTCAT
TCTTCTTGACTTCAAATTAATATTGTCAAACCTATTTTACTTACTTCTACCTTTTTTTA
TATCAAAAGTTTCCACCAAGAAAGAAATTCATATTATGATGCCAAGATTGGTTTGCATT
TGGGGTTGAACACATTGTAATCTCTTACGACCACATAATCAAGTGGTCTCTCTTTTTT
TGCTGTCTAA

>G2059 Amino Acid Sequence (conserved domain in AA coordinates:184-254)

MEDQFPKIETSMHDKLLSSGIYGLSSSTPPQLLGVPILFEGMKSPLLPASSTPSYFVS
PHDHELTSSIIHSPVASVPWNFLSEFPQSQHPDHHPSKPPNLTFLKEPKLLELSQSESN
MSPYHKYIPNSFYQSDQNRNEWVEINKLTINYPSKGFNYWLSSTTKTQPMKSKTRKVVQT
TPTKLYRGVRQRHWGKWVAEIRLPRNRTRVWLGTFFETAEQAAMAYDTAAYILRGEFAHL
NFPDLKHQLKSGSLRCMIASLLESKIQQISSQVSNSPPPPKVGTPEQKNHMKMESG
EDVMMKKQKSHKEVMEGDGVQLSRMPSLMDLIWDALSFPHSS*

>G2105 (42..1487)

CTCTCTGACTTGAACCTCTTCTCTCTACCGAATCAAACCAATGGAGGATCATCAAACC
ATCCACAGTACGGTATAGAACACCATCTTCTCAATTCTCTCTGATCTCTTCGGCTTCA
ACCTCGTTTCAGCGCCGGACCAGCACCATCGTCTTCATTTCACCGACCATGAGATAAGTT
TATTGCCACGTGGAATACAAGGGCTTACGGTGGCTGGAAACAACAGTAACACTATTACAA
CGATCCAGAGTGGTGGCTGTGTGGTGGGTTTAGTGGCTTTACGGACGGCGGAGGAACAG
GGAGGTGGCCGAGGCAAGAGACGTTGATGTTGTTGGAGGTCAGATCTCGTCTTGATCACA
AGTTCAAAGAAGCTAATCAAAGGGTCTCTCTGGGATGAAGTTTCTAGGATTATGTCGG
AGGAACATGGATACACTAGGAGTGGCAAGAAGTGTAGAGAGAAGTTCGAGAATCTCTACA
AGTACTATAAAAAACAAAAGAAGGCAATCCGGTCCGCGACAAGATGGTAAAACTATA
GATTTTTCCGGCAGCTTGAAGCGATATACGGCGAATCCAAAGACTCGGTTTCTTGCTATA
ACAACACGCACTTCATAATGACCAATGCTCTTCATAGTAATTTCCGCGCTTCTAACATTC
ATAACATCGTCCCTCATCATCAGAAATCCCTTGATGACCAATACCAATACTCAAAGTCAAA
GCCTTAGCATTTCTAACAAATTTCAACTCTCTCTCGATTGGATCTAACTTCTCTCTCTG
AAGGAAACGAACTACTAAAAGAGAGGGGATGCATTGGAAGGAAAAGATCAAGGAATTCA

TTGGTGTTCATATGGAGAGGTTGATAGAGAAGCAAGATTTTGGCTTGAGAAGTTGATGA
AGATTGTGGAAGACAAAGAACATCAAAGGATGCTGAGAGAAGAGGAATGGAGAAGGATTG
AAGCGGAAAGGATCGATAAGGAACGTTCTGTTTGGACAAAAGAGAGGGAGAGGATTGAAG
CTCGGGATGTTGCGGTGATTAATGCCTTGACGACTTGACGGGAAGGGCATTGATAAGGC
CGGATTCTTCGTCTCCTACAGAGAGGATTAATGGGAATGGAAGCGATAAAATGATGGCTG
ATAATGAATTTGCTGATGAAGGAAATAAGGGCAAGATGGATAAAAAACAAATGAATAAGA
AAAGGAAGGAGAAATGGTCAAGCCACGGAGGGAATCATCCAAGAACCAGAGAATATGA
TGATATACAACAATCAAGAACTAAGATTAAATGATTTTGTGAGATGATGACCAATGCC
ATCATGAAGGTTACTCACCTTCAAACCTCCAAGAACGCAGGAACCTCCGAGCTGCAGCAATG
CCATGGCAGCTAGTACAAAGTGCCTTTCATTGCTTGAAGGAGAAGGAGATCAGAACTTGT
GGGAGGTTTATGGTTTGAAGCAAAGGAAAGAAAATAATCATCAGTAAGCTACATTTTCA
TCTCTAAAATGAAGAATAAGAGAACTTAGAAACGAT

>G2105 Amino Acid Sequence (domain in AA coordinates: 100-153)
MEDHQNHPOYGIEQSSQFSSDLFGFNLVSPDQHHRLHFTDHEISLLPRGIQGLTVAGN
NSNTITTIQSGGCVGGFSGFTDGGGTGRWPRQETLMLLEVRSLDHKFKEANQKGPLWDE
VSRIMSEEHGYTRSGKKCREKFENLYKYYKTKEGKSGRRQDGNRYRFRQLEAIYGESK
DSVSCYNNQTQFIMTNALHSNFRASNIHNIVPHHQNPMTNTNTQSQSLISNNFNSSDL
DLTSSSEGNETTKREGMHWKEKIKFIVHMERLIEKQDFWLEKLMKIVEDKEHQRLRE
EEWRRIEAEERIDKERSFWTKERERIEARDVAVINALQYLTLGRALIRPDSSSPTERINGNG
SDKMMADNEFADEGNKGMDDKQMNKKRKEKWSSHGNNHPRTKENMMIYNNQETKINDFC
RDDDQCHHEGYSPSNSKNAGTPSCSNAMAASSTKCFPLLEGEQDQNLWEGYGLKQQRKENNH
Q*

>G2117 (49..465)
ATACTGTGCAACAAAAATTTCTTAAAGAACGCATAACTGTTTTTTTCATGGCTGGTTCT
GTCTATAACCTTCCAAGTCAAAACCCCTAATCCACAGTCTTTATTCCAAATCTTTGTTGAT
CGAGTACCACTTTCAAACCTGCTGCCACGTCAGACGACTCTAGCCGGACTGCAGAAGAT
AATGAGAGGAAGCGGAGAAGGAAGGTATCGAACCGCGAGTCAGCTCGGAGATCGCGTATG
CGGAAACAGCGTCACATCGAAGAACTGTGGTCCATGCTTGTTCAACTCATCAATAAGAAC
AAATCTCTAGTCGATGAGCTAAGCCAAGCCAGGGAATGTTACGAGAAGGTTATAGAAGAG
AACATGAACTTCGAGAGGAAAACTCCAAGTCGAGGAAGATGATTGGTGAGATCGGGCTT
AATAGGTTTCTTAGCGTAGAGGCCGATCAGATCTGGACCTTCTAATCGTCTCGTAAGCTT
GTTGGTTTTTTGTTGTTTATTAAAG

>G2117 Amino Acid Sequence (conserved domain in AA coordinates:46-106)
MAGSVYNLPQNPNPQSLPQIFVDRVPLSNLPATSDSSRTAEDNERKRRRKVSNRESAR
RSRMKQRHMEELWSMLVLQLINKNKS LVDLSQARECYEKVIEENMKLREENSKSRKMIG
EIGLNRFLSVEADQIWF*

>G2124 (87..923)
GAACAGCAAAAACCTAGATTTCCTGTTCAAGCTCAAGACCGTACAAAACCTTTGGAATCA
TATATAAAGATCTCGAGAATAGCATTATGAATATCGTCTCTTGGAAGATGCAAACGACG
AAGTTGCAGGCGGCGCTACGACAAGACGTGAAAGAGAAGTAAAAGAGGATCAAGAAGAAA
CCGAAGTCAGAGCCACCAGTGGCAAAACCGTAATTAAGAGCAGCCTACATCGATCTCTT
CTTCTTCTTCTCTGTTGATGAAATCCAAGGATCCGAGGATTGTTAGGGTTTCACGCGCT
TTGGAGGCAAGACCGTCAAGCAAGTGTGTACGTTACGTGGACTACGTGACAGACGCG
TGAGATTATCAGTCCCAACGGCTATTACGCTCTACGATCTTCAAGAACGGCTCGGTGTG
ACCAGCCTAGCAAAAGCCGTTGACTGGTTGCTTGATGCAGCTAAAGAGGAGATCGACGAGC
TACCTCCGTTACCTATCTCGCCGGAAAATTTACGATCTTCAACCATCATCAGTCTTCT
TGAATCTTGGTCAACGGCCCGGTCAAGATCCGACCAACTCGGGTTTAAAATCAATGGAT
GTGTACAAAAGTCTACTACTACTAGCCGCGAAGAAAACGATAGAGAGAAAGGAGAAAACG
ATGTCGTTTTACAAACAATCATCATGTTGGGTCTTATGGAACCTATCACAACCTGGAAC
ATCATCATCATCATCAACATTTGAGTTTACAGGCAGATTATCATAGTCATCAACTAC
ATAGTCTTGTCCCATTTCCATCACAATTTTGGTATGTCCAATGACGACATCAACACAA
CTACAATATACAATCTTTGTTTCCATCATCATCGTCAGCTGGTTTCAGGGACTATGGAGA
CATTAGATCCGAGGCAATGTAGCAACAATGGTGGTAGAGACATTGATAATCGGATGTCG
TCGGTCCAATTAACCGAACTAATAGCACTACAACGGCTAACATGTCGAGGCATCTAGGC
TCGGAGCGTTGTACAAGTAGAGGAAGTGATCACCATATGTGAAGTTAGATTATTGAAACG
ATATAATTGTTGTTTGTGTTTCAGAAATAAGGGGACAC

>G2124 Amino Acid Sequence (domain in AA coordinates: 75-132)

MNIVSWKDANDEVAGGATRRREREVKEDQEETEVRATSGKTVIKKQPTSISSSSSSWMKS
KDPRIVRVSRAFGGKDRHSKVCTLRGLRDRRVRLSVPTAIQLYDLQERLQVDQPSKAVDW
LLDAAKEEIDELPPLPI SPENFSIFNHHQSFLNLGQRPGQDPTQLGFKINGCVQKSTTTS
REENDREKGENDVVYTNHHVGSYGTYNLEHHHHHHQHLSLQADYHSHQLHSLVPFSPQ
ILVCPMTTSPTTTTIQSLFPSSSSAGSGTMTETLDPQM*

>G2140 (148..1254)

ACTCTCTTAACCTTTCGTTCTTCTCTCTACCTTCTTTTACCAACCTTCTCTTCTCTTACA
CACATATATATATACATATATAGAGAGAGAGAAGAGGACAAAGAGTTGAAAGATGAAGAC
TCTCATGTCTTCATAGAAACAAGTGATATGTGCGCTAAGAAAGAAGAAGAAGAAGAA
GAAGAAGACAGTTCTGAAGCCATGAACAACATACAAAATTACCAAAATGACCTCTTCTTT
CACCACATCATCTCTCATCATCACCATCATCATCATGATCCTTCTCAATCTGAAACTTTG
GGAGCATCCGGTAACGTTGGATCTGGTTTCACTATCTTCTCTCAAGATTCCGCTCTCTCCA
ATATGGTCTCTACCTCCACCTACCTCGATCCAACCACCATTTGATCAGTTTCTCTCTCTCT
TCTTCTTCTCCAGCATCTTCTACGGAAGTTTCTTCAACAGAAGTCGAGCTCATCATCAG
GGATTACAGTTTGGGTACGAGGGTTTGGTGGAGCCAGTCAGCAGCACATCATCATCAT
GAACAACCTCGGATCTTGTCTGGAAGCTTTAGGTCCGGTAGTACAAGCCGGGTCCGGTCT
TTTGGGTTACAAGCTGAGTTAGGGAAGATGACAGCACAAGAGATCATGGACGCTAAAGCT
TTGGCTGCTTCAAAGAGTCATAGTGAAGCTGAGAGAAGAAGAAGAGAGAGAATCAATAAT
CATCTCGCTAAGCTCCGTAGCATATTACCCAACACCACCAAAACGGATAAAGCGTCGTTA
CTAGCTGAAGTGATGATGATGAAAGAGTTGAAGAGAGAGACTTCAGTGATCTCAGAG
ACAAATCTTGTCCTCAACGGAAGCGATGAGTTAACGGTAGCTTTCACGGAGGAGGAAGAA
ACCGGAGATGGCAGATTTGTAATTAAGCGTCGCTTTGCTGTGAAGACAGGTCCGATCTC
TTGCCTGACATGATTAACATTGAAAGCTATGCGTCTCAAACGCTCAAGCGGAGATA
ACCACCGTTGGGGGACGAGTCAAGAACGTTTTGTTTGTACCGGAGAAGAGACTCCGGT
GAGGAAGTGAGGAAGAGTACTGTATAGGGACGATTGAGGAAGCTTTGAAAGCGGTGATG
GAGAAGAGCAATGTAGAGGAATCATCTTCTTCTGGAATGCTAAGAGACAGAGAATGAGT
AGTCACAACACTATCACTATCGTCAACAACAACAATATAATCAGAGGTAATCAATT
TTTTACTTAAATCGCTTTTTTTTCTTACTTTTCGGTGTATCTACTACGTGTGTGTGTGT
GGTTATGGAATGAATGTTGTACGTACGTTATACTATAGATATATGTGTGTGTGTGTGT
ATGTATAACGGAAGTATTTGTATCCGTTGTGGTCTTGGACTTTTGGTTTGGTTCTAAGAT
ACTTATTTTTTAAACCTTGTATCGTTGAGTTGGTTTTCTAGATATGCTTAATGGGAGTAT
GTGACGAAAAAAA

>G2140 Amino Acid Sequence (domain in AA coordinates:167-242)
MCAKKEEEEEEDSSSEAMNNIQNYQNDLFFHQLISHHHHHHHHDPQSSETLGASGNVGSF
FTIFSQDSVSPISLPPPTS IQPPFDQFPSPSSSPASFYGSFNRRAHHQGLQFGYEGF
GGATSAHHHHEQLRILSEALGPVVQAGSGPFLQALGKMTAQEIMDAKALAASKSHSE
AERRRRERINNHLAKLRSILPNTTKTDKASLLAEVIQHVKEKRETSVISETNLVPTESD
ELTVAFTEEEETGDRFVIKASLCCEDRSDLLPDMIKTLKAMRLKTLKABITTVGGRVKN
VLFVTGEESSEEBEYICGTIEEALKAVMEKS NVEESSSSGNAKRQRMSSHNTITIVE
QQQQYNQR*

>G2144 (102..1241)

ATTAGGGTTTTGTGTCGTGAGATTTGATTACACAAATTGCTGAATTTGGTTTCGATTAT
TGGTGTATTGTTTTCAAGATTTCCAGTGAGTTTCCGTTTATGGATCTGACTGGAGGAT
TTGGAGCTAGATCCGGCGGTGTGGACCGTGCCGGAACCAATAGGCCTTGAATCGCTAC
ATCTCGGTGACGAATTTCCGCAACTAGTGACGACTTTACCTCCCGAGAACCCCGCGGTT
CGTTCACGGCTTTGCTTGAGCTTCCACCTACACAAGCAGTGAGCTTCTCCATTTCACTG
ATTCTTCGTCTTCTCAACAAGCGGCAGTGACAGGGATCGGTGGAGAGATTCTCCGCCGC
TTCACTCTTTCCGGTGGGACATGGCTTTTCTTCTAAGTCACTCATGGAGCGAGCAG
CTCGTTTCTCGGTGATTGCCACTGAGCAACAAACCGAAATATCTCCGGGAGACTCCGA
CGAGCTCTGTACCTTCAATCAAGTGCTAATCTCGACAGAGTCAAGACGGAGCCTGCTG
AGACCGATTCTCTCAGCGTTGATTTCTGATTACGCGATTGAGAATCAAATCCCTTGCC
CTAACCAGAACAAATCGAAATGGGAAGAGGAAAGATTTGAAAAGAAGGGTAAAAGCTCGA
CGAAGAAGAACAAGCTCTGAAGAGAACGAGAAGCTGCCATATGTTACGTTAGAGCTC
GTCGTGGTCAAGCAACCGATAGCCATAGCTTAGCAGAACGAGCAAGAAGAGAGAAGATAA
ATGCACGAATGAAGCTGTTACAGGAACCTGGTCCAGGCTGTGATAAGATTCAAGGTACCG
CGCTGGTGTGATGAATCATTAACCATGTCCAGTCATTACAACGTCAGTGGAGATGC
TATCAATGAGACTTGCTGCGGTAAACCCGAGAATCGACTTCAATCTCGACACCATATTGG

CTTCAGAAAACGGTTCCTTAAATGGATGGGAGCTTCAATGCCGCACCAATGCAGCTTGCTT
GGCCTCAGCAAGCCATTGAGACCGAACAGTCCTTTCATCACCGGCACTGCAACAACCAC
CAACACAACAATGGCCTTTTGACGGCTTGAACCAGCCGGTATGGGGAAGAGAAGAGGATC
AAGCTCATGGCAATGATAACAGCAATTTGATGGCAGTTTCTGAAAATGTAATGGTGGCTT
CTGCTAATTTGCACCCAAATCAGGTCAAAATGGAGCTGTAAGTTGGGAAAACGGTAGAGA
TCATGAATGTGTATATACATCGTATAAGCTCGTTTCTCTCTATATAAATATAATCATAAA
TATAGATATCTGTTAAGAAGGTATCAGTCATTTGATTTCAGAGAGACAACACTGGTATGAT
TGTTTCTTATTCTTGACCAGATTTGACAATGTAGAATTTAGTAGGATATGATCATTTT
GATCTCGTTATATATA

>G2144 Amino Acid Sequence (domain in AA coordinates:203-283)
MDLTGGFGARSGGVGPCREPIGLES LHLGDEFRLVTTLPENPGGSFTALLELPPTQAV
ELLHFTDSSSSQQAAVTGIGGEIPPLHSFGGTLAPPSNSVLMERARF SVIATEQQNGN
ISGETPTSSVPSNSSANLDRVKTEPAETDSSQRLISDSAIENQIPCPNQNNRNGKRKDFE
KKGKSSTKKNKSSEENEKLPYVHVRRRQATDSHSLAERARREKINARMKLLQELVPGC
DKIQGTALVDEIINHVS LQRQVEMLSMRLA AVNPRIDFNLDTILASENGSLMDGSFNA
APMQLAWPQQAIETE QSFHHRQLQPPPTQQWPF DGLNQPVWGREEDQAHGNDNSNLMAVS
ENVMVASANLHPNQVKMEL*

>G2431 (47..1057)
CCCTTTTCGTTTATTATTTAAATTTCTTGGGTGCTTTC'TTAAATTTGTATGTGTTTATTAAT
GGAGATCAACAATAATGCCAACAATACTAATACTACTATTGATAATCACAAGGC AAAGAT
GAGCCTTGTTGTGTAACGGATGCTAAGCCAAGGTTGAAATGGACTTGATGATCTTCATCA
CAAATTCATCGAAGCCGTTAATCAACTTGGAGGACCTAACAAAGCAACACCTAAGGGTTT
GATGAAGGTTATGAGAGATCTCGGCTTACCTTATACCATCTCAAGAGCCATTTACAGAA
ATATCGGTTAGGGAAGAGCATGAAGTTCGATGATAACAAGCTAGAAGTTTCTCTGCATC
AGAGAATCAAGAAGTTGAGAGTAAAAACGATTCAAGAGATCTCCGAGGCTGCAGTGTAC
CGAAGAAAACAGCAATCCAGCTAAAGAAGGGCTACAAATCACAGAGGCTTTACAAATGCA
GATGGAAGTTCAAGAAGAACTTCATGAACAAATCGAAGTTCAGAGGCATTTGCAGGTGAA
GATTGAGGCACAAGGAAAGTATCTACAGTCCGTTTAAATGAAAGCTCAACAAACTCTCGC
TGGCTACTCATCTTCAAATCTCGGCATGGATTTTGCAGAGACCGAGCTCTCTAGATTAGC
TTCAATGGTTAAGCAGAGGCTGTCCAAGCACTTCGTTTCTCAGAGCTAACGCAAGTAGAAGA
AGAAGAAGAAGGTTTCTTGTGGTACAAGAAACAGAAAACAGAGGAATTAGTCAGCTGAG
ATGTTTCAGTAGAGAGCTCGTTGACATCTTCAGAGACCTCAGAGACAAAACCTGGATACTGA
CAATAACCTTAATAAATCGATTGAACCTCCGTTGATGGAGATCAACTCGGAAGTGATGAA
GGGGAAGAAGAGAAGCATAAACGACGTCGTTTGCCTGGAGCAGCCTCTAATGAAGAGAC
TTTTGGAGTTGATGATGATGAGCATTGGAAGTTGAGTTTGAATACTTACAAGAAAGACAT
GGAGGCGTGTACGAACATAGGACTAGGTTTAATTA AAAAAAAAAAACATTTTACTAAAGTT
ATATAAAAATGTTTTTAAAGAATCCA

>G2431 Amino Acid Sequence (conserved domain in AA coordinates:38-88)
MCLLMEINNNANNTNTTIDNHKAKMSLV LSTDAKPR LKWTCDLHHKFIEAVNQLGGPNKA
TPKGLMKVMEIPGLTLYHLKSHLQKYRLGKSMKFDDNKLEVSSASENQEVESKNDSRDLR
GCSVTEENS NPAKEGLQITEALQM QMEVQKKLHEQIEVQRHLQVKIEAQGKYLSVLMKA
QQTLAGYSSSNLGMDFARTELSRLASMVNRGCPSTSFSELTQVEEEEGFLWYKKPENRG
ISQLRCSVESLTSSETSETKLD TDNNLNKSIELPLMEINSEVMKGKKRSINDVVCVEQP
LMKRAFGVDDDEHLKLSLNTYKKDMEACTNIGLGFN*

>G2465 (86..1150)
CAATATTCTTTCTCCATTGAGATTAAGCTTCTTCTCGCTGTCGTCTCTCTATAGATCTT
GGTTCTTAGTCCCTTTTGAATAATAATGATGGTGGAGATGATTACGCTAAGAAAATGCA
GAAATGT CATGAATACGTTGAAGCACTTGAAGAAGAACAGAAGAAAATCCAAGTCTTTCA
ACGCGAGCTTCCTTTATGTTTAGAGCTTGTCAC TCAAGCGATCGAAGCTTGTCGGAAGGA
GTTATCTG GTACGACGCAACTACATCAGAACAGTGTT CAGAACAGACCACAAGTGTGTTG
TGGTGGTCCCTGTCTTTGAAGAGTTTATTCTTATCAAGAAAATTAGTTCTTGTGTGAAGA
AGTACAAGAAGAAGAAGAAGATGGTGAACATGAATCTTCTCCAGAACTTGTAATAA
TAAGAAATCAGATTGGCTTAGATCTGTT CAGCTATGGAATCATT CACCGGATCTAAATCC
AAAAGAGGAGCGGTG TAGCTAAGAAAGCGAAAGTGGTGGAGGTGAAACCAAAAGCGGTGC
GTTTCAGCCGTTTCAAAGCGCGTTTGGAGACTGATTGCAACCGGCGGTGAAAGTAGC
TAGTTTCGATGCCAGCGACGACGAGGTTCTACGACGGAACCTTGTTGGTGGTAAAAGTGA
TTTGATTAAAGCTGGAGATGAGGAAAGACGGATAGAGCAGCAGCAATCGCAGTCCGATAC

GCATAGAAAACAAAGGCGGTGCTGGTCGCCGAATTACACCGTCGATTCTCTAAACGCGCT
TCAGCAGCTTGGAGGATCTCATGTTGCTACACCAAAGCAAATCAGGGATCACATGAAGGT
TGATGGATTAAACAAACGACGAAGTTAAAAGCCATTTACAGAAATATAGACTTCACACAAG
AAGGCCAGCAGCAACATCCGTGGCGGCACAAAGTACCGGAATCAGCAACAACCACAATT
TGTGGTGGTTGGAGGCATATGGGTACCATCGTCACAAGATTTTCCACCACCGTCCGATGT
AGCCAACAAGGGTGGTGTATATGCTCCGGTTGCGGTGGCGCAATCTCCAAAACGTTCTGTT
GGAGAGAAGTTGCAACTCGCCGGCGGCATCTTCTCTACAAATACAAATACTTCTACTCC
TGTGTCATAATCTGATAGTCATACTATAATCATCTCTGATGTTGATTTTGGTGTAGGTT
TGAAAATGTTTATGTGAATGTAA

>G2465 Amino Acid Sequence (conserved domain in AA coordinates:219-269)

MMVEMDYAKMQKCHEYVEALEEEQKKIQVFQRELPLCLELVTQAI EACRKELSGTTTTT
SEQCSEQTTSVCGGPVFEEFIPIKKISSLCEEVQEEEEEDGEHESP ELVNNKSDWLRS
VQLWNHSPDLNPKEERVAKKAKVVEVKPKSGAFQPFQKRVLETDLQPAVKVASSMPATTT
SSTTETCGGKSDLIKAGDEERRIEQQQSQSHTRKQRRCWSP ELHRRFLNALQQLGGSHV
ATPKQIRDHMKVDGLTNDEVKSHLQKYRLHTRRPAATSVA AQSTGNQQQPQFVVVGGIWV
PSSQDFPPPSDVANKGGVYAPVAVAQSPKRSLSRSCNSPAASSSTNTNTSTPVS*

>G2583 (38..607)

CAAATCAGAAAATATAGAGTTTGAAGGAACTAAAAGATGGTACATTCGAGGAAGTTCCG
AGGTGTCCGCCAGCGACATGGGGTTCTTGGGTCTCTGAGATTCGCCATCCTCTATTGAA
GAGAAGAGTGTGGCTTGGAACTTTTGAACCGGCAGAGCGGTGCAAGAGCATACGACCA
AGCGGCTCTTCTAATGAACGCCAAACGCTAAGACCAATTTCCCTGTCTGTAATAATCAGA
GGAAGGCTCCGATCACGTTAAAGATGTTAACTCTCCGTGATGTCACCAAAGTCATTATC
TGAGCTTTTGAACGCTAAGCTAAGGAAGAGCTGCAAAGACCTAACGCCCTTCTTTGACGTG
TCTCCGTCTTGATACTGACAGTTCCACATTTGGAGTTTGGCAGAAACGGGCCGGGTGCGAA
AACAAGTCCGACTTGGGTGATGCGCCTCGAAGTTGGGAACGTAGTCAACGAAAGTCCGGT
TGACTTAGGGTTGACTACGATGAACAAACAAAACGTTGAGAAAGAAGAAGAAGAAGA
AGCTATTATTAGTGATGAGGATCAGTTAGCTATGGAGATGATCGAGGAGTTGCTGAATTG
GAGTTGACTTTTGACTTTAACTTGTGCAAGTCCACAAGGGGTAAGGGTTTTTC

>G2583 Amino Acid Sequence (domain in AA coordinates:4-71)

MVHSRKFRGVRQRQWGSWVSEIRHPLLRVWLGT FETAEEAARAYDQAALLMNGQNAKT
NFPVVKSEEGSDHVKDVNSPLMSPKSLSELLNAKLKSKDLTPSLTCLRLD TDSSHIGV
WQKRAGSKTSPTWVMRLELGNVVNESAVDLGLTTMNKQNVKEEEEEEEAII SDEDLAME
MIEELLNWS*

>G2724 (1..651)

ATGGAAATAGAAATAGGAGAGGTCCATGGACTGTGGAAGAAGACATGAAGCTCGTCAGT
TACATTTCTCTTACCGTGGAAGGAAGATGGAACTCCCTCTCTCGTTCTGCTGGACTGAAT
AGAACGGGGAAAAGTTGCAGATTGCGGTGGCTAAATTATCTCCGGCCGGATATCCGCCGT
GGAGACATATCCCTTCAAGAACAAATTTATCATCCTTGAAGTCCATTCTCGTTGGGGAAAT
CGGTGGTCAAAGATTGCTCAACATTTACCGGAAGAACAGATAACGAGATAAAGAATTAT
TGGAGAACACGTGTTCAAAGCATGCAAACTTCTAAAATGTGACGTGAACAGCAAGCAA
TTCAAAGACACCATCAAACATCTCTGGATGCCTCGTCTCATCGAGAGAATCGCCGCCACT
CAAAGTGTCAAATTTACCTCTAACCCTACTCGCCTGAGAACTCCAGCGTCGCCACCGCC
ACGTGATCAACGTCGTCTGAGGCTGTGAGATCGAGTTTCTACGGTGGTGATCAGGTG
GAATTTGGAACGTTGGATCATATGACAAATGGTGGTTATTGGTTCAACGGCGGAGATACG
TTTGAACTTTGTGTAGTTTGTGACGAGCTCAACAAGTGGCTCATACAGTAG

>G2724 Amino Acid Sequence (conserved domain in AA coordinates:7-113)

MEIEIRRGPTVEEDMKLVSYISLHGEGRWNSLSRSAGLNRTGKSCRLRWLNLYLRPDIRR
GDISLQEQFIIELEHSRWGNRWSKIAQHLPGRTDNEIKNYWRTRVQKHAKLLKCDVNSKQ
FKDTIKHLWMPRLIERIAATQSVQFTSNHYPENSSVATATSSSSSEAVRSSFYGGDQV
EFGTLDHMTNGGYWFNGGDTFETLCSFDELNKWLIQ*

>G377 (1..396)

atgggtctctcgcatTTTTCCAACAGCGTCAGAAGGAGTACTACCATTCTGGTGATGAAC
acgggtgttTCAATCACTCTGTTGAAGAACATGGTGAGGTCTGTTTTCAAATTGTTGCA
tccgagactgaatcttccatggagatagacgacgagcctgaagatgattttgttactaga
agaatctcgataacacagttcaagtctctatgtgagaacatagaagaggaagaagaagag
aaaggtgtggagtggtgtgtgtgcctttgtgggttaaagaggaagaggaagtgagtga
ttgggttcttgcaagcatttctccacagagcttgtctagacaactgggttgtaataac

cacaccacatgccctctttgcagggtccattctctag
>G377 Amino Acid Sequence (domain in AA coordinates:85-128)
MGLSHFPTASEGVLP LLVMNTVVSITLLKNMVRSVFQIVASETESSMEIDDEPEDDFVTR
RISITQFKSLCENIEEEEEKGV ECVCLCGFKEEEVSELVSKHFFHRA CLDNWFGNN
HTTCPLCRSIL*
>G428 (97..1032)
TTACTTTTGTGTTTCTTCATATTCTTCAGAAGCAAGCACAAAGGCTAGGGATCGAAGAAGC
GGCGATCACTGATCGTATCTCACTACGATCACATTAATGGATAGAATGTGTGGTTTCCGC
TCGACGGAAGACTATTTCGGAGAAAGCGACGTTGATGATGCCGTCGGATTATCAGTCTTTG
ATTTGTTCAACCACCGGAGACAATCAAAGACTGTTTGGATCCGACGAACTCGTACCGCT
TTGTCCTCGGAGTTGCTTCCGCGTATTTCGAAAAGCTGAGGATAATTTCTCTCTTAGTGTC
ATCAAATCCAAATCGCTTCTCATCTTTGTATCCTCGCTTACTCCAAACCTACATCGAT
TGCCAAAAGGTGGGAGCGCCTATGGAAATAGCGTGATATTGGAAGAGATTGAGCGAGAG
AACCATGTGTACAAGAGAGATGTTGCTCCATTATCTTGCTTTGGAGCTGATCCTGAGCTT
GATGAATTCATGGAACCTACTGTGATATATTGGTTAAATACAAACCGATCTTGCAGAGG
CCGTTTCGACGAGGCTACAACCTTTTATAAACAAGATTGAAATGCAGCTTCAGAACTTGTGC
ACTGGTCCAGCGTCTGTACAGCTCTTTCAGATGATGGTGCGGTTTCATCTGACGAGGAA
CTGAGAGAAGATGATGACATAGCAGCGGATGACAGCCAAACAAAGAAGCAATGACCGCGAT
CTGAAGGACCAGCTACTACGCAAAATTTGGTAGCCATATCAGTTTCATTGAAACTCGAGTTC
TCTAAAAGAAGAAGAAAGGGAAGCTACCAAGAGAAGCAAGACAAGCGTTGCTCGATTGG
TGGAAATGTTTATAATAATGGCCTTACCCTACTGAAGGCGACAAATAGCTCTGGCTGAA
GAAACAGGTTTGGATCAAAAACAATCAACAATTGGTTTATAAACCAAGGAAACGCCAT
TGGAAAGCCTTCGGAGAACATGCCGTTTGTATGATGGACGATTCTAATGAAACATTCTTT
ACCGAGGAATGAAAAGAGAGACATGGGATTGTGCATTGTATAATTTTACACTGTTTTCC
CAAGAAAAGAAAACAGTAAAAGCTTTTGGTAAATGGGACATCATCGGAATGAATGGAA
CCAGTTAGCCAAAACGGTCAAGGGCGTGGCGTAACGAGACATTGTATTGAAATAGTGGC
AATATTATGTCTAATCTTCCAATGGTCCAAAATGATAGATTTCTTATTGTATTGAAC
CTTACTTAGATAGCTGATGTGTCAACTAAATAATTTATTTTCATCCTTATACTACTTGTA
TCAATGTCTCTAATTGATCAATTGTTGCTTGCTATTCAAAAAAAAAAAAAAAAAAAAAA
>G428 Amino Acid Sequence (domain in AA coordinates: 229-292)
MDRMCGFRSTEDYSEKATLMMPSDYQSLICSTTGDNQLRFGSDELATALSSELLPRIRKA
EDNFSLSVIKSKIASHPLYPRLLQTYIDCQKVGAPMEIACILEEIQRENHVKRDVAPLS
CFGADPELDEFMETYCDILVKYKTDLARPFD EATTFINKIEMQLQNLCTGPASATALSDD
GAVSSDEELREDDIAADDSQQRSDNRDLKDQLLRKFGSHISSLKLEFSKKKKKGKLPRE
ARQALLDWNVHNKWPYPTEGDKIALAETGLDQKQINNWF INQRKRHWKPS ENMPFDM
DDSNETFFTEE*
>G447 (241..3501)
CTTTTAAAGAGCTTAAAAATTTGCTTTGAAGCTTCAAATATTCTTATGAACTAAAAAGAA
GAAAAAGCTTTTGTCTTTTTCCTTAGCAGCAGAATGATTTTGTCTTCCAAAATTATT
ACTATTTAGTTTCTCTCGTGCTCTTCTCTTGAGCAAATACAGATTGCTTAATTTTGCTGA
AGAAGAAGAACTCTGTTTCTTCCCTGCACCAAACCAATTTTTCGTTCTTTCTATAAACC
ATGAAAGCTCCATCAAATGGATTTCTTCCAAGTTCCAACGAAGGAGAGAAGAAGCCAATC
AATTCCTCAACTATGGCAGCTTGTGCAGGGCCTTTAGTTTCATTACCTCCTGTGGGAAGT
CTTGTGTTTACTTCCCTCAAGGACACAGCGAGCAAGTTGCAGCATCGATGCAGAAGCAA
ACAGATTTTATACCAAATTACCCAAATCTTCTTCTAAGCTGATTGCTTGCTTCACAGT
GTGAATAAGTATGATAGAGAAGCATTTGCTAGCTTCTGATATGGGCTTGAAGCTAAACAGA
CAACCTACTGAGTTTCTTGAAGACTCTTACTGCAAGTGACACAAGCACTCATGGTGGA
TTCTCTGTACCGCGTCGTGCAGCTGAGAAAATATTCCTCCTCTTGATTTCTCGATGCAA
CCGCTGCGCAAGAGATTGTAGCTAAAGATTTACATGATACTACATGGACTTTTCAGACAT
ATCTATCGAGGCCAACCAAAAAGACACTTGCTTACCACAGGTTGGAGCGTTTTTGTAGC
ACAAAGAGACTATTTGCGGGTGATTGAGTTTGTGTTGTAAGAGATGAGAAATCACAGCTG
ATGTTGGGTATAAGACGTGCAATAGACAAACTCCGACTCTTTCCTCATCGGTATATCC
AGCGACAGTATGCACATTGGGATACTTGCAGCTGCAGCTCATGCTAATGCCAATAGTAGC
CCTTTTACCATCTTCTTCAATCCAAGGGCAAGTCCTTCAGAGTTTGTAGTTCTTTAGCC
AAATACAACAAAGCCTTATACGCTCAAGTATCTCTAGGAATGAGATTCCGGATGATGTT
GAGACTGAGGATTGTGGGGTTCGTAGATATATGGGTACAGTCACAGGTATTAGTGATCTT

GACCTGTAAAGATGGAAAGGCTCACAATGGCGTAATCTTCAGGTAGGATGGGATGAATCA
 ACAGCTGGAGATAGGCCAAGCCGAGTATCCATATGGGAAATCGAACCCTCATAACTCCT
 TTTTACATATGTCCTCCTCCATTTTTCAGACCTAAGTACCCGAGGCAACCCGGGATGCCA
 GATGATGAGTTAGACATGGAAAATGCTTTCAAAGAGCAATGCCTTGGATGGGAGAAGAC
 TTTGGGATGAAGGACGCACAGAGTTCGATGTTCCCTGGTTTAACTCTAGTTCAATGGATG
 AGTATGCAGCAAAACAATCCATTGTTCAGGTTCTGCTACTCCTCAGCTCCCGTCCGCGCTC
 TCATCTTTTAACTACCAAACAATTTGCTTCCAACGACCCCTTCCAAGCTGTTGAACTTC
 CAATCCCCAAACCTCTCTTCGCAAAATCCCAATTCAACAACCGAACACGGTTAACCAT
 ATCAGCCAACAGATGCAAGCACAACCAGCCATGGTGAAATCTCAACAACAACAACAACA
 CAACAACAACAACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACA
 TCACAGCAACAGGTGCAGCAACAAGGGATTTATAACAATGGTACGATTGCTGTTGCTAAC
 CAAGTCTCTTGTCAAAGTCCAAACCAACCTACTGGATTCTCTCAGTCTCAGCTTCAGCAG
 CAGTCAATGCTCCCTACTGGTGCTAAAATGACACACCAGAACAATAAATCTATGGGGAAT
 AAAGGCTTGTCTCAAATGACATCGTTTGGCGCAAGAAATGCAGTTTCAGCAGCAACTGGAA
 ATGCATAACAGTAGCTAGCTTATTAAAGAAACAGCAAGAAGTCTCTCTCCATTCTATTA
 CAACAAAATCTGTCCCAAAATCCTCAGCAACTCCAAATGCAACAACAATCATCAAAACCA
 AGTCTTTCAACAACAGCTTCAGTTGCAGCTACTGCAGAAGCTACAGCAGCAGCAACAGCAG
 CAGTCGATTCTCCAGTAAGCTCATCCTTACAGCCACAATTATCAGCGTTGCAGCAGACA
 CAAAGCCATCAATTGCAACAACCTTCTGTCTCTCAAATCAACAGCCCTTGGCACATGGT
 AATAACAGCTTCCCAGCTTCAACTTTTCATGCAGCCTCCACAGATTGAGGTGAGTCTCAG
 CAGCAAGGACAGATGAGTAACAAAATCTTGTAGCCGCTGGAAGATCACATTCTGGCCAC
 ACAGATGGAGAAGCTCTTCTTGTTCAAACCTCACCTTCCGCCAATAACACGGGACATGAT
 AATGTTTACCGACAAATTTCTGTAGCAGAAATCAACAGCAAGGACAAGCTGCATCTGTA
 TCTGCATCTGATTAGTCTTTGAGCGCGCAAGCAATCCGGTCCAAGAGCTTTATACAAAA
 ACTGAGAGCCGGATCAGTCAAGGCATGATGAATATGAAGAGTGTGGTGAACATTTCAGA
 TTTAAAAGCGCGGTAAACAGATCAAAATCGATGTATCCACAGCGGAACGACGTAAGTCTCCT
 GATGTTGTTGGCCCTGTACAGCAGCAACAACTTTCCCACTACCATCATTTGGTTTTGAT
 GGAGACTGCCAATCTCATCATCAAGAAACAACCTTAGCTTTCCCTGGTAATCTCGAAGCC
 GTAACCTTCTGATCCACTCTATTCTCAAAGGACTTTCAAACCTTGGTTCCCAACTATGGC
 AACACACCAAGAGACATTGAGACGGAGCTGTCCAGTGTGCAATCAGTTCTCAGTCATTT
 GGTATTCCCAGCATTTCCCTTTAAGCCCGGATGTTCAAATGAGGTTGGCGGCATCAATGAT
 TCAGGAATCATGAATGGTGGAGGACTGTGGCCCAATCAGACTCAACGAATGCGAACATAT
 ACAAAGGTTCAAAAACGAGGGTCAGTAGGTAGATCAATAGATGTTACCCGTTATAGCGGC
 TATGATGAACCTTAGGCATGACTTAGCGAGAATGTTTGGCATCGAAGGACAGCTCGAAGAT
 CCGCTAACCTCTGATTGGAAACTCGTCTACACCGATCAGGAAAACGATATTTTACTAGTT
 GGTGATGATCCTTGGGAAGAGTTTGTGAAGTGCCTGCAGAACATAAAGATACTATCATCA
 GTAGAAGTTAGCAAAATGAGCTTAGACGGAGATCTTGCAGCTATCCCAACCACAAACCAA
 GCCTGCAGCGAAACAGACAGCGGAAATGCTTGGAAAGTACACTATGAAGACACTTCTGCT
 GCAGCTTCTTTCAACAGATAGAAATAAAAAGATGCAAAATATACCAAGTCAACTTACATTA
 TCATTCCAGGCCATCGCAAAGTACATGTTTTTTTTTGTGTGTATGTACTGCAAAACAACA
 ACTGAGAAGAAGATACTGCACGGTATATAAACATTTTTTATAGGACAGTGATTGATT
 TTTTCACTTAACTTGATGTTGTTGTACTTTCTTGTTCATATTTGTATAACAAGTATAA
 TGCTTGACAAGTCTATGAGGAGCATATCTTATACAGAGATACTAAGATGTAATGTTAATG
 TAACTAAACAATTACCTTCATTAATCATGAATCCTTTGGTCGTTTAAAA

>G447 Amino Acid Sequence (conserved domain in AA coordinates:22-356)

MKAPSNGLPSSNEGEKKPINSQWLHACAGPLVSLPPVGSVLVYFPQGHSEQVAASMQKQ
 TDFIPNYPNLPKSLICLLHVSLLHADTETDEVYAQMTLQPVNKYDREALLASDMGLKLN
 QPTEFFCKTLTASDSTHGGFSVPRRAAEKIFPPLDFSMQPPAQEIVAKDLHDTTWTFRH
 IYRGQPKRHLLTTGWSVFTVKRLFAGDSVLFVRDEKSQLMLGIRNRANRQTPTLSSSVIS
 SDSMHIGILAAAAHANANSSPFTTIFNPRASPSEFVVPLAKYNKALYAQVSLGMRFRMMF
 ETEDCGVRRYMGTVTGISDLDPVRWKGSRWNLQVGVDESTAGDRPSRVSIWEIEPVITP
 FYICPPFFFRPKYPRQPGMPDDELDMENAFKRAMPWMGEDFGMKDAQSSMPGLSLVQWM
 SMQQNNPLSGSATPQLPSALSSFNLPNNFASNDPSKLLNFQSPNLSSANSQFNKPNTVNH
 ISQQMQAQPAMVKSQQQQQQQQQHQHQQQQLQQQQQLQMSQQQVQQQGIYNNGTIAVAN
 QVSCQSPNQPTGFSQSQLQQQSMPLTGAKMTHQNINSMGNKGLSQMTSFAQEMQFQQQLE
 MHNSSQLLRNQEQSSLSLHSLQNLQNPQQLQMQQSSKPSPSQQLQLQLQLQLQQQQQ
 QSIPPVSSSLQPLSALQQTQSHQLQQLLSQNNQQLAHGNNSFPASTFMQPPQIQVSPQ

QQGQMSNKNLVAAGRSHSGHTDGEAPSCSTSPSANNTGHDNVSPTNFLSRNQQQGQAASV
SASDSVFERASNPVQELYTKTESRISQGMNMKSAGEHFRKSAVTDQIDVSTAGTTYCP
DVVGPVQQQQTFFPLPSFGFDGDCQSHHPRNNLAFPGNLEAVTSDPLYSQKDFQNLVPNYG
NTPRDIETELSSAAISSQSGFIPSIPIFKPGCSNEVGGINDSGIMNGGGLWPNQTRMRTY
TKVQKRGSVGRSIDVTRYSGYDELRLHDLARMFGIEGQLEDPLTSDWKLVYTDHENDILLV
GDDPWEEFVNCVQNIKILSSVEVQQMSLDGDLAAIPTTNQACSETDSGNAWKVHYEDTSA
AASFNR*

>G464 (41..760)

CTCTGCTGGTATCATTGGAGTCTAGGGTTTTGTTATTGACATGCGTGGTGTGTCAGAATT
GGAGGTGGGAAGAGTAATCTTCCGGCGGAGAGTGAGCTGGAATTGGGATTAGGGCTCAG
CCTCGGTGGTGGCGCTGGAAAGAGCGTGGGAGGATTCTTACTGCTAAGGATTTCTCTTC
CGTTGGGTCTAAACGCTCTGCTGAATCTTCTCTCACCAAGGAGCTTCTCCTCCTCGTTC
AAGTCAAGTGGTAGGATGGCCACCAATTGGGTTACACAGGATGAACAGTTTGGTTAATAA
CCAAGCTATGAAGGCAGCAAGAGCGGAAGAAGGAGACGGGGAGAAGAAAGTTGTGAAGAA
TGATGAGCTCAAAGATGTGTCAATGAAGGTGAATCCGAAAGTTCAGGGCTTAGGGTTTGT
TAAGGTGAATATGGATGGAGTTGGTATAGGCAGAAAAGTGGATATGAGAGCTCATTCTGTC
TTACGAAAACCTTGGCTCAGACGCTTGAGGAAATGTTCTTTGGAATGACAGGTACTACTTG
TCGAGAAAAGGTTAAACCTTTAAGGCTTTTAGATGGATCATCAGACTTTGTACTCACTTA
TGAAGATAAGGAAGGGGATTGGATGCTTGTGGAGATGTTCCATGGAGAATGTTTATCAA
CTCGGTGAAAAGGCTTCGGATCATGGGAACCTCAGAAGCTAGTGGACTAGCTCCAAGACG
TCAAGAGCAGAAGGATAGACAAAGAAACAACCCTGTTTAGCTTCCCTTCCAAAGCTGGCA
TTGTTTATGATATTGTTTGGAGTTTGAATTTACTCGATACTTTTGAAGAAAGTATTTTG
GAGAATATGGATAAAAGCATGCAGAAGCTTAGATATGATTGAAATCCGGTTTTTCGGATAT
GGTTTTGCTTAGGTCAATTCGTAGTTTTCAGTTTGTCTTCTTTGGCTGTGTAC
CAATTATCTATGTTCTGTGAGAGAAAGCTCTT

>G464 Amino Acid Sequence (domain in AA coordinates: 20-28, 71-82, 126-142, 187-224)

MRGVSELEVKSNLPAESELELGLSLGGAWKERGRILTAKDFPSVGSKRSAESSSHQ
GASPPRSSQVVGWPPIGLHRMNSLVNNQAMKAARAEEDGEEKVVKNDKDVSMKVNPK
VQGLGFVKVNMMDGVGIGRKVDMRAHSSYENLAQTLEEMFFGMTGTTCREKVKPLRLLDGS
SDFVLTYEDKEGDWMLVGDVPRMFINSVKRLRIMGTSEASGLAPRQEQKDRQRNNPV*

>G557 (192..698)

CAGAGATCTGACGGCGGTAGCAGAGTAATCTATTCCTTCCCAAATGTCTCGCAATTAGA
TTCTTTCCAAGTTCTTCTGTAAATCCCAAGTCCCGCTCTTTTCTCTTTATCCTTTTCAC
CAGCTTCGCTACTAAGACAACAAATCTTTCCCTCTCTCTCGCCTGATCGATCTTCAAA
GAGTAAGAAAAATGCAGGAACAAGCGACTAGCTCTTTAGCTGCAAGCTCTTTACCATCAA
GCAGCGAGAGGTCATCAAGCTCTGCTCCACATTTGGAGATCAAAGAAGGAATTGAAAGCG
ATGAGGAGATACGGCGAGTGCCGGAGTTTGGAGGAGAAGCTGTGCGGAAAAGAACTTCGG
GTAGAGAATCTGGATCGGCGACCGGTGAGGAGCGGACACAGGCGACTGTGCGGAGAAAGTC
AAAGGAAGCGAGGGAGGACACCGGCGGAGAAAGAGAACAAGCGGCTGAAGAGGTTGTTGA
GGAACAGAGTTTCAGCTCAGCAAGCAAGAGAGAGGAAAAAGGCTTACTTGAGCGAGTTGG
AAAACAGAGTGAAGACTTGGAGAACAAAACTCTGAACCTGAAGAGCGACTCTCTACTC
TTCAGAACGAGAACAGATGCTTAGACATATTCTGAAGAACACAACAGGAAACAAGAGAG
GAGGTGGTGGTGGTTCTAATGCTGATGCAAGCCTTTGATCTCTTCTTCTTCTTGTGTTA
TATTTTGTGGATAAAATTTACAGAGAATTGTATCAATAATTATCATGTTAAATTTATAT
GGGATGTGAGAGCTAATATTGCAATTGTAGACCAAGTTCTTTAAAAA

AAAAA

>G557 Amino Acid Sequence (domain in AA coordinates: 90-150)
MQEQATSSLAASSLPSSSSERSSSAPHLEIKEGIESDEEIRRVPEFGGEAVGKETSGRES
GSATGQERTQATVGESQRKRGRTPAEKENKRLKRLLRNRVSAQQARERKKAYLSELENRV
KDLENKNSLEERLSTLQENENQMLRHILKNTTGNKRGGGGGSNADASL*

>G577 (44..2155)

AAAAACAGACTGAGAGAGAGAGAGAGAGTGTGTTGTTGGCCATGGGATGCACGGCCTC
CAAGCTCGACAGTGAGGATGCTGTCGCTGCAAGGAGCGGCGCGCTCTTATGAAGGA
CGCCGTCTACGCTCGTCAACATCTCGCCGCGCTCACTCTGACTACTGCGCTCCCTTCG
TCTACTGGCTCTGCCCTCTCCTCTTCGCGCGCGGCGAGCCCCTCTCCGTCTCCGAGAA
TACTCCGCTGTTTTCTCGCCCTTCTCCAGTCAGGACGCGCCACGTGTCCCTTCTTC

CCATTCCCCAGAACCCCTCCTCCGCCATCCGACGCAAGCCTAAGCCTACTAGGCCTAG
 GAGGCTTCCACACATTCTCTCCGACTCCTCTCCTTCTTCTCTCTGACCAGTTTCTA
 TCCCCTGCTCACCAGAACTCTACTTACTCTCGCTCTCCATCTCAAGCTTCTCTGTCTG
 GAACTGGGAGAATTTCTACCCCTCCTCTCCCCCGACTCCGAGTACTTCGAACGCAAAGC
 TCGCCAGAACCACAAGCACCGTCTCCTTCCGACTACGACGCCGAAACTGAAAGATCCGA
 CCACGATTACTGCCACTCACGGAGAGATGCCGCCGAGGAAGTTCACTGCAGCGAGTGGGG
 CGACGACCACGACCGTTTCACTGCCACCTCTTCGTCCGACGGAGATGGGGAGGTGCAAAAC
 TCACGTTTCCAGATCCGGTATTGAAGAAGAGCCTGTGAAACAACACATCAAGACCCAAA
 TGAGCAAAGAGCACTCTGACCATGTTACCACTTCTTCCGACTGCTACAAGACCAAATTGGT
 GGTAAAGGCACAAGAATTTGAAGGAGATCCTTGACGCCGTTCAAGACTACTTCGACAAGGC
 TGCCTCCGCTGGGGACCAGGTCTCCGCCATGCTTGAGATCGGCCGGGCTGAGCTCGACCG
 CAGCTTCAGCAAGCTGAGGAAGACGGTGTATCATCAAGCAGTGTGTTTCAGCAACTTGAG
 CGCAAGCTGGACCTCAAAACCCCATTTGGCAGTCAAATACAAGCTCGATGCATCTACCTT
 GAATGATGAACAAGCGGCCCTCAAGAGCCTCTGCTCCACTCTAGACCGACTCCTCGCTTG
 GGAAAAGAAGCTTTATGAGGATGTCAAGGCAAGAGAAGGAGTTAAGATTGAGCAGAGAA
 GAAGCTGTCTGCGCTGACCATGAGTATAGGAGTATAAGGGAGGTGATGAATCCAAGCTAGACAA
 GACTAAAACCTTCCATAACAGACTGCAATCACTCATCATTGTTTCTTTCAGAAGCTGTTTT
 AACCACGTCTAATGCCATTCTCCGCCCTCCGGGACACTGACCTTGTCCCTCAGCTTGTGTA
 ACTCTGCCACGGATTATGTACATGTGGAAGTCAATGCACGAGTATCACGAAATCCAGAA
 CAACATCGTGCAACAAGTCCGTGGCCTGATCAACCAAACAGAGAGAGGTGAGTCAACATC
 AGAGGTACACCGGCAGGTGACGCCGGACCTAGAGTCAGCTGTGTCTTGTGGCATTTCGAG
 CTTCTGTGCGCATCATTAAATTCCAGAGGGAGTTTATGCTCTCTCCACGCATGGTTCAA
 GCTGAGCCTGGTTCCCTGAGCAACGGAGACCCAAAGAAACAGCGGCCAGACTCATTGTC
 CTTGTGCGAGGAGTGAAGCAGAGCCTGGAACGGGTGCCTGACACAGTGGCGTCAGAAGC
 CATAAAGAGCTTTGTAAACGTGGTACATGTGATATCAATAAAGCAGGCGGAAGAGGTGAA
 GATGAAGAAACGCACGGAGAGTGCAGGAAAGGAGCTGGAGAAGAAAGCATCTCACTGAG
 GAGCATAGAGAGGAAGTACTACCAGGCATACTCGACGGTTGGGATAGGCCCTGGACCGGA
 GGTGTTGGACTCACGGGACCCGCTATCTGAGAAGAAATGTGAGCTGGCGGCATGTGAGAG
 GCAGGTGGAGGATGAGGTAATGAGGCACGTGAAGGCTGTGGAGGTGACACGAGCTATGAC
 TCTCAACAATCTACAAACCGGCCTGCCCAATGTATTCAGGCCCTTGACCAGCTTCTCATC
 TCTCTTCACTGAATCTCTCCAGACTGTCTGTTCTCGTTCTTCTACTCCATCAACTGATTATG
 TCCAAGTTTCTCATTTATTTTAAAGCTCTCATTACGTTGGTATCATGTAAATTTGAGGAT
 TGATTAAATTGAGTCTTGTGGTTTGTGAGGACTCACAATCTTCTCATTAAAAA
 AAAAAA

>G577 Amino Acid Sequence (domain in AA coordinates: TBD)
 MGCTASKLDSSEDAVRCKERRRLMKDAVYARHHLAAHSDYCRSLRLTGSALSSFAAGEP
 LSVSENPVAVFLRPSSSQDAPRVPSSHSPPEPPPPPIRSKPKPTRPRRLPHILSDSSPSSS
 PATSFYPTAHQNSTYSRSPSQASSVWNWENFYPPSPDSEYFERKARQNHKRPSPDYDA
 ETERSDDHYCHSRDAEEVHCSEWGDDHDFRFTATSSSDGDGEVETHVSRSGIEEPEVKQ
 PHQDPNGKEHSDHVTSSDCYKTKLVVRHKNLKEILDAVQDYFDKAASAGDQVSAMLEIG
 RAELDRSFSKLRKTVYHSSSVFNSLSASWTSKPPLAVKYKLDASTLNDEQGLKSLCSTL
 DRLLAWEKKLYEDVKAREGVKIEHEKKLSALQSQYKGGDESKLDKTKTSITRLQSLIIV
 SSEAVLTSSNAILRLRDLVLPQLVELCHGLMYMWSMHEYHEIQNNIVQQVRGLINQTE
 RGESTSEVHRQVTRDLESASVSLWHSSFCRIIKFQREFICSLHAWFKLSLVPLSNGDPKKQ
 RPDSFALCEEWKQSLERVPTVASEAIKSFVNVVHVVISIKQAEVVKMKRTESAGKELEK
 KASSLRISIERKYYQAYSTVGIGPGEVLDSRDPLSEKKCELAACQQRQVEDEVMRHVKA
 VTRAMTLNLTGLPNVVFQALTSFSSLFTESLQTVCSRSYSIN*

>G674 (1..786)-

ATGGTGTTTAAATCAGAAAAATCAAACCGGGAAATGAAATCAAAGGAGAAGCAAAGGAAG
 GGATTATGGTCACCCGAGGAAGATGAGAAGCTTAGGAGTCATGTCTCAAATATGGCCAT
 GGATGCTGGAGTACTATTCTCTTCAAGCTGGATTGCAGAGGAATGGGAAGAGTTGTAGA
 TTAAGGTGGGTTAATTATTTAAGACCTGGACTTAAGAAGTCTTTATTCATAAACAAGAG
 GAAACTATACTTCTTTCACTTCATCCATGTTGGGTAACAAATGGTCTCAGATATCGAAA
 TTCTTACCAGGAAGAACCACAACGAGATCAAAAACTATTGGCATTCTAATCTAAAGAAG
 GGTGTAACCTTTGAAACAACATGAAACCACAAAAAATCAACACCTTTAATCACAAC
 TCACTTGAGGCCTTGACAGATTCAACTGAAAGATCTTCTCATCTATCAATGTCCGAGAA
 ACGTCTAATGCTCAAACCTCAAGCTTTTCGCCAAATCTCGTGTTCTCGGAATGGTTAGAT

CATAGTTTGCTTATGGATCAGTCACCTCAAAAGTCTAGCTATGTTCAAAATCTTGTTTTTA
CCGGAAGAGAGAGGATTTCATTGGACCATGTGGCCCTCGTTATTTGGGAAACGACTCTTTG
CCTGATTTTCGTGCCAAATTCAGAAATTTTGTGGATGATGAGATATCATCTGAGATCGAG
TTCTGTACTTCATTTTCAGACAACTTTTGTTCGATGGTCTCATCAACGAGCTACGACCA
ATGTAA

>G674 Amino Acid Sequence (domain in AA coordinates: 20-120)
MVFKEKSNREMKSKQKRLWSPEEDEKLRSVLYKHGHCWSTIPLQAGLQRNGKSCR
LRWVNYLRPGLKKSLFTKQEETILLSLHSM LGNKWSQISKFLPGRTDNEIKNYWHSNLKK
GVTLKQHETTKKHQTPLITNSLEALQSSTERSSSSINVGETSNAQTSSFSPLNVFSEWLD
HSLMDQSPQKSSYVQNLVLPBERGFIGPCGPRYLGNDSLDFVPNSEFLLDDEISSEIE
FCTSFSDNFLFDGLINELRPM*

>G736 (1..513)
ATGGCGACTCAAGATTCTCAAGGGATTAACTCTTTGGCAAACTATTGCATTTAACACT
CGAACATAAAAAATGAAGAAGAGACACACCCGCCGAGCAAGAAGCCACAATAGCCGTT
AGATCATCATCATCATCGGATCTGACGGCCGAGAAGCGTCCGATAAGATCATAGCATGT
CCAAGATGCAAGAGCATGGAGACAAAGTTCTGTTACTTCAACAATAACAACGTAATCAG
CCTCGACACTTTTGTAAAGGCTGCCACCGTTACTGGACCGCGGTGGTGCCTCCGGAAC
GTTCCCGTCGGCGCCGGTCTGTCGGAAGTCCAAACCACCTGGTCTGTCGTGGTGGTATG
CTTGGAGATGGAATGGTGTTCGCCAAGTCGAGCTTATAAATGGCTTGCTCGTTGAGGAG
TGGCAGCATGCCGAGCCGAGCTCACGGTAGTTTCCGGCATGATTTCCCATGAAGCGG
CTCCGGTGTACTCCGACGGTCAATCGTGCTGA

>G736 Amino Acid Sequence (domain in AA coordinates: 54-111)
MATQDSQGIKLFKGTIAFNTRTIKNEEETHPPEQATI AVRSSSSDLTAEKRPDKIIAC
PRCKSMETKFCYFNNGNNGPRHFCKGCHRYWTAGGALRNVPVGAGRRKSKPPGRVVVGM
LGDGNGVRQVELINGLLVEEWQHAAAAHGSFRHDFPMKRLRCYSDGQSC*

>G903 (96..1496)
CCCGGGTCGACCCACGCTCCGCTCTCTCTCTGAACATACAAAAACCTACTTTTAAT
TTCTCTTCCAAGAAGTCAAGAACCAGAGAAGACATGACAAGTGAAGTTCTTCAAACAA
TCTCAAGTGGATCAGGTTTGTCTCAGCCACAGAGCTCATCAACCTGGATCATGATGAAT
CTCTCATCAATCCTCTCTTGTGAAGAAAAAGAGAAATCTCCCTGGAAATCCTGATCCGG
AAGCTGAAGTGATAGCTTTATCCCCACGACCTTGATGGCTACGAACCGGTTCTCTATGTG
AGGTATGTGGCAAAAGTTTCCAAGAGAGACCAAACTTACAGCTTCAATCGGCGAGGACATA
ATCTTCCATGGAAGTTGAAGCAGAGGACAAGCAAAGAAGTGAGAAAAAGTGTCTACGTTT
GCCCCGAGAAGACATGTGTCCACCATCACTCTCTAGAGCTCTAGGCGATCTCACTGGAA
TCAAAAAGCATTTTGTCCGGAACACGGGGAGAAGAGTGGACGTGCGAGAAATGTGCTA
AGAGATACGAGTCCAATCTGATTGGAAGCTCATTTCCAAGACTTGTGGTACTAGAGAGT
ACCGTTGCGATTGTGGCACCATTCTTCAAGGCGAGACAGCTTTATCACTCATAGAGCTT
TCTGCGATGCTTACGCGAAGAAACCGCTAAGATAAACGCAGTGTCTCATCTCAACGGTT
TAGCCGCGGCTGGAGCCCGAGGATCAGTTAATCTCAACTATCAATATCTCATGGGAACAT
TCATCCCACCGCTTCAACCATTTGTACCACAACCGCAAAACAAATCCAAACCATCATCATC
AACATTTTCAGCCACCAACTTCTTCGTCGCTCTCTCTATGGATGGGACAAGATATCGCGC
CGCCTCAACCGCAACCGGACTACGATTGGGTTTTTGGAAACGCTAAGGCAGCGTCTGCTT
GCATTGATAATAATAATACTCACGATGAGCAGATTACGCAAAACGCAAAACGCAAGTTTGA
CCACTACCACTACTCTCTGCCCCCTTCTTATTACGAGCGACCAACCACAAAACGCAA
ACGCAAAATTCAAACGTGAATATGTCCGCGACAGCTTTACTACAGAAAGCTGCTGAAATTG
GCGCTACTTTTACAACAACCGCAGCGACCAATGACCCATCAACGTTTCTTCAAAGTTTCC
CGCTTAAATCCACCGATCAAACCAACAGTTATGACAGTGGCGAAAAGTTTTTGTCTTGT
TCGGGTCTAACAGAACATTGGGTTAATGAGTCGTAGTCATGATCATCAAGAGATCGAGA
ACGCTAGAAATGACGTTACGGTTGCGTCTGCTTGGATGAATTACAGAATTACCCTTGGGA
AACGTAGAAGAGTTGATGGTGGAGGTGAAGTGGGTGGAGGAGGGCAAATCGGGATTTC
TCGGGGTGGTGTACAAACGTTGTGCCATCCATCGTCTATCAATGGATGGATTGAAAGA
GTTTAAAAATTTCCGGGTTAATGCATAAATTACGTAAAAGAAGAAGGAATCTTTGTCTAT
TTCCACCATTTTCTAAGATAACATATGTATATGGTAATGAAGTTGTTTTCTTTTATTAA
TTCAATATTCTAAAACCTTATGATATATGTATAATGAATGTGTTTATCTTCAA

>G903 Amino Acid Sequence (domain in AA coordinates: 68-92)
MTSEVLQTISSGSGFAQPQSSSLDHDESLINPPLVKKRNLPGNPDPEAEVIALSPPTL
MATNRFLCEVCGKGFQDQNLQLHRRGHNL PWKLKQRTSKEVRKRVYVCPEKTCVHHSS

RALGDLTGIIKKHFCRKHGEKKWTCEKCAKRYAVQSDWKAHSKTCGTREYRCDCGTIFSR
DSFITHRAFCDALAEETAKINAVSHLNGLAAGAPGSVNLNYQYLMGTFFIPPLQPFVFPQ
QTNPNHHHQHFQPTSSSLSLWMGQDIAPPQPQPDYDWVFGNAKAASACIDNNNTHDEQI
TQANANASLT'TTTLSAPSLFSSDQPNANANSNVNMSATALLQKAAEIGATSTTTAATND
PSTFLQSFPLKSTDQTTSDYSGEKFFALFGSNNNIGLMSRSHDHQEIENARNDVTVASAL
DELQNYPPWKRRRVDGGGEVGGGQTRDFLGVGVQTLCHPSSINGWI*

>G917 (32..679)

TTAGGGT'TTTAGAAAGATAGATCGATTGAAGATGAGGAAAGGTAAGAGAGTGATAAAAAA
GATAGAGGAGAAAATAAAGAGACAAGTGACATTCGCAAAGAGAAAGAAGAGTCTAATCAA
GAAGGCATATGAACCTCTCTGTCTCTGCGATGTCCACCTTGGTCTCATCATCTTCTCTCA
CTCCAAACAGGCTCTACGATTTCTGCTCCAACCTCTACCAGCATGGAGAATCTCATCATGAG
ATACCAAAGGAAAAAGAAGGTCAAACCACTGCAGAACACAGTTTCCACTCGGATCAGTG
TTCAGATTGCGTGAAGACGAAGGAATCAATGATGAGAGAGATAGAGAATCTTAAGCTGAA
TCTTCAATTGTACGACGGACATGGCTTGAATCTCTTGACCTACGACGAGCTCCTTTCTTT
TGAGCTCCATCTCGAATCTTCTCTACAACATGCTCGAGCTCGCAAGTCTGAGTTCATGCA
TCAGCAGCAGCAGCAACAAACAGATCAAAAGCTTAAGGAAAAGAAAAGGGTCAAGGAAG
CTCTTGGGAGCAGCTGATGTGGCAAGCAGAGAGACAGATGATGACGTGTCAAAGACAAAA
AGATCCTGCGCGCGCAATGAAGGAGGAGTTCCTTTTTTACGGTGGGGAACAAACCCACCG
ACGTTCTTTCACCTCCTTAAGCTACCACAACCAAGGCCCAAATACAGGCCCATAACTTCTCT
CTATCTATAAAAAACAACCTGATAGTAAAAAGTATGACCCGTTTGGTTTCGTTATGTTG
ATACCAGACTATTAATTAACCTCGGTTAGACGTATTTACGACTTGATGCTATCTAGACCT
TTTTGCCCTTCAAAAAA

>G917 Amino Acid Sequence (conserved domain in AA coordinates:2-57)

MRKGRVKKIEEKIKRQVTFARRKSLIKKAYELSVLCDVHLGLIIFSHSNRLYDFCSN
STSMENLIMRYQKEKEGQTTAEHSFHSDDQSDCVKTKESMMREIENLKLNLQLYDGHGLN
LLTYDELLSFLHLESSLQHARARKSEFMHQQQQQQTDQKLKGKEKGQSSWEQLMWQAE
RQMMTCQRQKDPAPANEGGVFLRWGTTTHRRSSPP*

>G921 (116..1024)

CCAAGATCGACTCTTACTTCGAATCTCTCTCAACTTTCTTCTCAGCTTACGGGAACCTC
CACACATATACATCCACAAGAACCCATATCGAAGATTCTCTACATATATTTACATGGA
TCAGTACTCATCCTCTTTGGTCGATACTTCATTAGATCTCACTATTGGCGTTACTCGTAT
GCGAGTTGAAGAAGATCCACCGACAAGTGCTTTGGTGGAAGAATTAACCGAGTTAGTGC
TGAGAACAGAAGCTCTCGGAGATGCTAACTTTGATGTGTGACAACCTACAACGTCTTGAG
GAAGCAACTTATGAATATGTTAACAAGAGCAACATAACCGAGAGGGATCAAATCAGCCC
TCCCAAGAAACGCAATCCCGCGGAGAGAGGACGCATTACGCTGCGCGGTTATTGGCGG
AGTGTGCGGAGAGTAGCTCAACGGATCAAGATGAGTATTTGTGTAAGAAGCAGAGAGAAGA
GACTGTGCTGAAGGAGAAAGTCTCAAGGGTCTATTACAAGACCGAAGCTTCTGACACTAC
CCTCGTTGTGAAGAGATGGGTATCAATGGAGGAAATATGGACAGAAAGTACTAGAGACAA
TCCATCTCCAAGAGCTTACTTCAAATGTGCTTGTGCTCCAAGCTGTTCTGTCAAAAAGAA
GGTTCAGAGAAGTGTGGAGGATCAGTCCGTGTTAGTTGCAACTTATGAGGGTGAACACAA
CCATCCAATGCCATCGCAGATCGATTCAAACAATGGCTTAAACCGCCACATCTCTCATGG
TGGTTACAGCTTCAACACCCGTTGCAGCAAACAGAAGAAGTAGCTTGACTGTGCCGGTGAC
TACCGTAGATATGATGAATCGAAGAAAGTGACGAGCCCAACGTCAAGAATCGATTTTCC
CCAAGTTCAGAAACTTTTGGTGGAGCAAATGGCTTCTTCTTAAACCAAGATCCTAACTT
TACAGCAGCTTTAGCAGCAGCTGTACCGGAAAATGTATCAACAGAATCATACCGAGAA
ATAGTTTAGCTTCAAATCCGTTAGAGTTTTTAGATTGAAATTTGTCATGAGTAAGAGAA
AGAGAGTAGATTATAATCCNTTGTGATACTGAAAAAAAAAAAAAAAAAAAAA

>G921 Amino Acid Sequence (domain in AA coordinates: 146-203)

MDQYSSSLVDTSLDLTIGVTRMRVEEDPPTSALVEELNRVSAENKKLSEMLTLMCDNINV
LRKQLMEYVNKSNITERDQISPPKKRKSAPAREDAFSCAVIGGVSESSSTDQDEYLCKKQR
EETVVKKEKVSRYVYTEASDTTLVVKDGYQWRKYGQKVRDNPSPRAYFKCACAPSCSVK
KKVQRSVEDQSVLVATYEGEHNPMPQSIDSNNGLNRHISHGGSASTPVAANRRSSLTVP
VTTVDMIESKKVTSPTSRIDFPQVQKLLVEQMASSLTGDPNFTAALAAAVTGKLYQONHT
EK*

>G922 (1..1449)

ATGGTGGCTATGTTTCAAGAAGATAATGGAACATCTTCTGTAGCTTCATCACCCTTCAA
GTCTTCTCAACTATGTCACTCAACAGACCGACTCTCCTCGCTTCTTCATCTCCGTTTCAT

TGTCTCAAAGATCTCAAACCAGAGGAGCGTGGTCTCTACTTAATCCACCTCTTGCTAACT
TGTGCCAACCACGTGGCTTCAGGTAGCCTCCAAAACGCTAACGCAGCGCTCGAGCAGCTC
TCTCACCTCGCTTCTCCTGACGGCGACACGATGCAGCGAATCGCTGCTTACTTCACCGAA
GCGCTTGCCTAACAGAATCCTTAAGTCTTGGCTGGTCTTTACAAGGCTCTTAACGCAACT
CAGACAAGAAGCTAACAATGTCTCTGAGGAGATTATGTTAGAAAGACTCTTCTTTGAGATG
TTCCCCGATACTCAAAGTCTCTTACTTGCTCACTAATCGAGCTATACTCGAGGCTATGGAA
GGAGAGAAGATGGTTCATGTGATTGATCTCGATGCTTCTGAGCCAGCTCAATGGCTTGCT
TTGCTTCAAGCTTTTAACTCTAGGCCTGAAGGTCCACCTCATTTGAGAATCACTGGTGT
CATCACCAGAAGGAAGTGTCTGAACAAATGGCTCATAGACTCATTGAGGAAGCAGAGAAA
CTCGATATCCCGTTTCAGTTTAAATCCCGTTGTGAGTAGGTAGACTGTTTAAATGTAGAA
CAGTTGCGGGTTAAACAGGAGAGGCTTAGCCGTTAGCTCGGTTCTTCAATTGCATACC
TTCTTGGCCTCTGATGATCTCATGAGAAAGAACTGCGCTTTACGGTTTTCAGAACAAAC
CCTAGTGAGATTGACTTGACAGAGAGTTCTAATGATGAGCCATGGCTCTGCAGCTGAGGCA
CGTGAGAATGATATGAGTAACAACAATGGGTATAGCCCTAGCGGTGACTCGGCTCATCT
TTGCTTTTACCAAGTTTCAGGAAGGACTGATAGCTTCTCAATGCTATTTGGGGTTTGTCT
CCAAAGGTCATGGTGGTCACTGAGCAAGACTCAGACCACAACGGCTCCACACTAATGGAG
AGGCTATTAGAATCACTTTACACCTACGCAGCATTGTTTGATTGCTTGAAACAAAAGTT
CCAAGAACGCTCTCAAGATAGGATCAAAGTGGAGAAGATGCTCTTCGGGGAGGAGATCAAG
AACATCATATCTCGAGGGATTGAGAGAAGAGAAAGACACGAGAAGCTTGAGAAATGG
AGCCAGAGGATCGATTGGCTGGTTTGGGAATGTTCTCTTAGCTATTATGCGATGTTG
CAGGCTAGGAGATTGCTTCAAGGGTGGGTTTGTATGGGTATAGAATCAAGGAAGAGAGC
GGGTGCGCAGTAATTTGCTGGCAAGATCGACCTCTATACTCGGTATCAGCTTGAGATGC
AGGAAGTGA

>G922 Amino Acid Sequence (conserved domain in AA coordinates:225-242)

MVAMFQEDNGTSSVASSPLQVFSTMSLNRPDLLASSSPFHCLKDLKPEERGLYLIHLLLT
CANHVASGSLQANANAALQLSHLASPDGDTMQRIAAYFTEALANRILKSWPGLYKALNAT
QTRTNVSEEIHVRRLLFFEMFPIKVSYLLTNRAILEAMEGEKMHVHIDLDASEPAQWLA
LLQAFNSRPEGPPLHRLITGVHHQKEVLEQMAHRLIEAEKLDIPFQFNPVVSRLDCLNVE
QLRVKTGEALAVSSVLQLHTFLASDDLMRKNCALRFQNNPSGVLDLQRLVLMMSHGSAAEA
RENDMSNNNGYSPSGDSASSLPLPSSGRTDSFLNAIWGLSPKVMVTEQSDSHNGSTLME
RLLESLYTYAALFDCLTKVPRTSQDRIKVEKMLFGEEIKNIISCEGFERRERHEKLEKW
SRIDLAFGNVPLSYAMQLQARRLLQCGFDGYRIKEESGCAVICWQDRPLYSVSAWRC
RK*

>G932 (206..1213)

CCACGCGTCCGACCACTTGTACCTCTTGTCTTAAGTACTCTTTAACCTACAATTCCT
AAGCTCTCAAGCCACAAAAAACCACAAACCGTTCTTACCAATATATATATCTGATCATC
ATCAAAGTCCTTCTCTCTGCTCATACCACAAACCGTTCCATTCTTCCCCTAATCACAAG
TGATATTACATAGAGAAGATAGAGATGGGAAGACCACCATGCTGTGACAAGATTGGAGT
GAAGAAAGGACCATGGACACCAGAGGAAGATATCATCTTGGTTTCTTACATCCAAGAACA
TGGTCCTGGAAACTGGAGATCTGTGCCTACTCACACAGGTTTGAGGAGATGTAGCAAAAG
CTGTAGATTGAGGTGGACTAATTATCTTCGACCTGGGATCAAGCGTGGAAATTTACCCGA
GCATGAAGAGAAGATGATCTCCATCTTCAAGCTCTTTGGGAAACAGGTGGGCAGCTAT
AGCATCATATCTTCCAGAAAGGACAGACAATGATATAAAGAACTATTGGAACACTCATTT
GAAGAAAAAGCTCAAGAAGATGAATGATTCTTGTGATAGTACTATCAACAATGGCCTTGA
TAATAAAGACTTCTCCATATCAAACAAAAACACTACCTCACATCAAAGCAGCAACTCCAG
TAAAGGTCAATGGGAGAGAAGGCTTCAGACAGATATCAACATGGCTAAACAAGCTCTTGT
TGATGCCCTTGTCTATTGACAAACCAACCAAAACCACTAATTTTTCTATTCCCGATCTTGG
TTATGGTCCATCATCTTCTCGTCTCTACCACCACCACCACCACCACCACCACGAG
AAACACTAATCCATACCCATCTGGGGTCTATGCTTCAAGTGCTGAGAACATTGCTCGTTT
GCTTCAGAAATTTATGAAGACACACCAAGACCTCGGTGCCCTTGCCGTTGACGCCAC
CGAGATGGCTATCACCACGGCAGCTTCGAGCCCTAGCAACCGAAGGAGACGGAGAAGG
GATTGACCAATCTTTGTTTCAGCTTCAACTCCATAGATGAAGCTGAAGAGAAGCCTAACT
AATAGACCATGACATTAATGGTCTAATTACACAAGGCTCTCTTCTTTGTTTCGAGAAATG
GCTCTTTGATGAGCAAGGCCACGATATGATCATCAATAACATGTCATAGAGGGTCAGGA
AGTGTGTTGTTAGAAAGCATTAAAGTTTGACGATTTGCTTGAGGAACCACGAGGCTTAGT
TATAACAATTTGTATAATTAAGTACTCTTAGTTTTGTTTTCAATCCTTATTATGATCA
TATTGCAGTAATTAGGGATTTTAGTCTTAGTAGTAACTCTTAAGTTTTAACACATTTTT

CTCTATCTTTTATAGTAGTAACTCTTTATTTTTCCTTAAATCTTTGTCGACGTGGAGATG
ATATCTTCTATGTAGTAGAACTCAAAGTGTACATCATCTTTATTAATGTAACGTCTTT
TAAAAAAAAAAAAAAAAA

>G932 Amino Acid Sequence (domain in AA coordinates: 12-118)
MGRPPCCDKIGVKKGPWTPPEEDIILVSYIQEHGPGNWRSPVHTGLRRCSSKSCRLRWITNY
LRPGIKRGNFTEHEEKMILHLQALLGNRWAAIASYLPERTDNDIKNYWNTHLKKLKKMN
DSCDSTINGLDNKDFSISNKNNTTSHQSSNSSKGQWERRLQTDINMAKQALCDALSIDKP
QNPTNFSIPDLGYGPSSSSSSSTTTTTTTTTTRNTNPYPGVSASSAENIARLLQNFMKDT
PKTSVPLPVAATEMAITTAASSPSTTEGDGEGIDHSLFSFNSIDEAEKPKLIDHDINGL
ITQGSLSLFEKWLFDQSHDMIINNMSLEGQEVLF*

>G599 (152..1579)

TCGACAGAACAGCTTCGTTGTCACTTGTCACTTCTATAAATCGCATCCCCATTGACAACTT
TTCATTCCATCAAACTCTCTCTCTATATCTCTCTCTCTATATCTCTCTCTATATCT
CTCTCTCTCTCACTCTCTCTTTCTTTCAAATGGAAAACTCATGGTTCCGACATGGAG
ACCCGACCCGGTTTACCGTCCACCGGAAACACCACTCGAACCGATGGAGTTTTAGCTCG
TTTCATGGAGCGTCTCTGCTCTCGAAGTCTCAAGGCTCTAACACCAACCAACCTCAGAT
TCTCTCTCTCAAAACCGAAGAAGAAGAAGAAGAAGAACCCATCTCTCTGTCTAGACGG
CGACGGCGACACGGAAGACACCGGACTTGTACCGGAAACCCATTCTCCTTCTGCTTGTTC
AGAACTTCTCAAATGGTTCATGGATCGTATCTTGTCTCACTCTCAAGAAGTATCACCAAG
AACATCTGGTCCGGCTATCTCACAGTAGTGGTCCACTTAATGGTTCTTTGACCGACAGTCC
TCCTGTGTCTCTCCGAATCCGACGACATTAAGCAATTTTGCAGAGCGAACAATAATTC
ATTGAACAGTGTAAATTTCTCAGTTCCGTTCAACGGCGGCAACTCCGGGACCTATAACCGC
TAGAGCTACACAGTCCAAAGACGGTGGGACGGTGGCTTAAGGACCGGAGAGAGAAAAAGAA
AGAGGAGACTCGGGCTCATAACGCTCAGATTACGCTGCTGTCTCTGTCTCGCCGGCTTGC
TGCAGCTGTGCTGCTATTGTCAGCAGCCACCGCTGCGTCTTCTAGCTGTGGTAAGGATGA
GCAGATGGCTAAAACTGACATGGCCGTTGCTTCTGTCTGCGACCCCTTGTGGCTGCTCAGTG
TGTGGAAGCTGCTGAAGTTATGGGAGCTGAGAGAGAGTATTTGGCTTCTGTTGTTAGCTC
CGCCGTCAATGTTCTGCTTCTGCGGAGATATTATGACTCTCACCGCCGGAGCAGCTACAGC
TTTAAGAGGAGTGCAACATTTGAAGGCAAGGGCAATGAAGGAAGTGTGGAACATAGCATC
AGTGATACCAATGGATAAAGGACTCACTTCTACAGGAGGAAGCAGCAATAATGTTAATGG
TAGCAATGGAAGCTCAAGCAGTAGTCACAGTGGTGAACCTGTACAAACAGGAGAATTTCTT
GGGAACCTGTAGTAGAGAAATGGCTCGCTAGAGGTTGTGAACCTCTCAAACGCACTCGCAA
AGGTGATCTCCACTGGAAGATAGTATCTGTTTACATCAACAAAATGAATCAGGTTATGTT
GAAGATGAAGAGCAGGCATGTTGGAGGAACCTTCACCAAGAAGAAAAAGAACATTGTGCT
TGATGTGATCAAGAATGTCCCGGCTTGGCTGGACGACATTTGCTAGAGGGAGGAGATGA
TLTAAGATACCTTCGGTTTGAAGACGGTTATGCGAGGTGATGTTGAATTCGAGGTCAAGAG
CCAAAGGAATATGAAATGTGGACACAAGGTGTCTCAAGGCTTCTGTTCTTGCTGCTGA
GAGGAAGTTTAGGATGTGAATAAACGTTCAATGGCTGCTTGGTTTAAGTGTGAGTTTTTT
TTTAACCTATGTGGTCAAATTTTATTAGTAGGGGTTCTTTTAAGGTAATGGTTTTTTGGG
TTGGGTATAGGATAAAATGGACCTACCAAGTCAAGGTGAGGAAGCATTTGGGTAAACAAAA
CTTAGTGGGGGTGATCTGTAATATCTATGTTCTTAGTTTTTTTTTGGTTGTTGGTGGTCT
TTTTGTATAAAAAACAAAGTTGAAGTAATAGATATATAGTATGTTTTAATTTTAAA

>G599 Amino Acid Sequence (domain in AA coordinates: 187-219, 264-300)

MEKLMVPTWRPDPVYRPPETPLEPMEFLARSWSVSALEVSKALTPPNPQILLSKTEEEEE
EPISSVVDGDDTETDGLVTGNPFSFACSETSQMVMRILSHSQEVSPRTSGRLSHSSG
PLNGSLTDSPPVSPPEDDIKQFCRANKNSLNSVNSQFRSTAATPGPITATATQSKTVGR
WLKDRREKKKEETRAHNAQIHAASVAVAGVAAAVAAIAAATAASSSCGKDEQMAKTDMAVA
SAATLVAAQCVEAAEVMGAEREYLASVVSSAVNVRSAGDIMTLTLAGAATALRGVQTLKAR
AMKEVWNIA SVIPMDKGLTSTGGSSNNVNGSNGSSSSSHSGELVQQENFLGTCSREWLR
GCELLKTRKGDLDHWKIVSVYINKMNQVLMKMSRHVGGFTFKKKKNIVLDVIKNVPAPW
GRHLLGGDDLRYFGLKTVMRGDVEFEVKSQREYEMWTQGVSRLLVLAERKFRM*

>G804 (114..1139)

ATACTCAAGAATTTATAGGTTATAAGTAAAAATTCAGTACAAGTTGTTTGTGTTGTTA
TTCCATTTTCTTGTGTGTTTTTTTCCCATTAATTTATAAATTTTATAAGCAATATGGAGT
CCCACAACAACACAGAGCAACAACACCACTGGTTCCGGCCCATCTGGTCCCATCCA
TGGGACCAATCTCCGGTTCAGTCTCATTAAACCACTGCTCCAACTCCACTACCACCA
CCGTACCGCCGCTAAACACCCGCAAAACGACCGTCCAAGGACCGTCACATCAAAGTAG

ACGGACGTGGCCGGAGGATACGTATGCCGGCTATCTGCGCAGCACGTGTCTTCCAATAA
CACGTGAGTTACAACACAAATCGGACGGCGAGACTATAGAGTGGCTGCTCCAACAAGCGG
AGCCAGCTATCATCGCAGCCACCGGAACCTGGAACCATACCGGCGAATATCTCTACTTTGA
ACATCTCTCTTCGAAGCAGTGGCTCTACTCTTTTCAGCTCCACTGTCTAAATCTTTCCACA
TGGGAAGAGCGGCTCAAAACGCTGCCGTTTTTGGGTTCAGCAACAGCTTTATCATCCTC
ATCATATCACGACAGATTCTTCTTCTTCTTCTTCTTCCAAAACATTCCGTGAAGAAGATC
TTTTTAAAGATCCTAATTTTCTAGATCAAGAACCGGTTCAAGATCACCTAAACCGGGAT
CCGAAGCTCCTGATCAAGATCCGGGTTGACCCGGTCAAGAACACAAAATATGATACCGC
CGATGTGGGCACTAGCGCCAACGCCAGCCTCCACAAACGGAGGTAGTGCTTTTTGGATGT
TACCAGTCGGAGGAGGAGGAGGTCCGGCTAACGTTTCAGGATCCATCACAGCACATGTGGG
CGTTTAAATCCGGGTCATTACCCGGGTCGAATCGGGTCGGTTTCAGCTAGGGTCTATGTAG
TGGGAGGTCAACAGTTAGGGTTAGGTGTTCAGAAAATAACAATTTGGGGCTATTTCCG
GCGGAGGAGGAGACGGTGGTCCGGTTCGGTCTCGGAATGAGTCTTGAGCAAAAGCCTCAAC
ATCAAGTGAGTGATCATGCTACTAGAGACCAAAATCCTACTATAGATGGTTCTCCTTGAA
AGACTTCATGATTTCTTTGGTTTTTAAAAAGTGTGAATGTGTGATTTATTGCAACTTTTG
TTGAGGACTCCAATGTTAATATGGGTTTTAGGGTTGGCTTTTCGGGATTGCCAAATTGTT
ATT

>G804 Amino Acid Sequence (domain in AA coordinates: 54-117)
MESHNNNQSNNTTGSALVPSMGPISGSVSLTTTAPNSTTTTVAATPAKRPSKDRHI
KVDGRGRRIRMPAICARVFQLTRELOHKSDGETIEWLLQQAEPALIAATGTGTIPANIS
TLNISLRSSGSLAPLSKSFHMGRAAQNAAVFGFQQQLYHPHHITDSSSSSLPKTFRE
EDLFKDPNFDQEPGSRSPKPGSEAPDQDPGSTRSRTQNMIPPMWALAPTASTNGGSAF
WMLPVGGGGGPANVDPSQHMWAFNPGHYPRIGSVQLGSMVLVGGQQLGLGVAENNNLGL
FSGGGDGGRVGLGMSLEQPKQHVSDHATRDQNPTIDGSP*

>G1062 (297..1781)

CAAAAAAAAAAGTTTCAATTTTGAAGCTCTGAGAAATGAAATCTATCATTCTCTCTCTC
TATCTCTATCTTCTTTTCAGATTTCGTTCTTCAATTTCATGAAATCCTCGTGATTCTAC
TTTAATGCTTCTCTTTTTTACTTTTCCAAGTCTCTGAATATCAAAGTATATATCTTTT
GTTTTCAAACCTTTTGCAAAATGTCTTCAAGCTTCCAAATTCAGTTAAAGGTCTCAACT
TTGCAAAATTTCTCTAAAGGTTTCAGACTTTGGGGTAAAGGTGTCAACTTTGGCGATGG
GTCTTGACGGAAACAATGGTGGAGGGGTTGGTTAAACGGTGGTGGTGGAGAAAGGAAG
AGAACCAGGAAGGTTTCATGGGAAGGAATCAAGAAGATGGTTCTTCTCAGTTTAAGCCTA
TGCTTGAAGGTGATTGGTTTTAGTAGTAACCAACCATCCACAAGATCTTCAGATGTTAC
AGAATCAGCCAGATTTCAGATACTTTGGTGGTTTTCTTTTAAACCTAATGATAATCTTC
TTCTTCAACACTCTATTGATCTTCTTCTTCTTGTCTCTTCTCAAGCTTTTAGTCTTG
ACCTTCTCAGCAAAATCAGTTCTGTCAACTAACAACAACAGGGTTGTCTTCTCAATG
TTCTTCTTCTGCAAAACCTTTTGATAATGCTTTTGAGTTTGGCTCTGAATCTGGTTTTC
TTAACCAAATCCATGCTCCTATTTCGATGGGGTTTGGTTCTTTGACACAATTGGGGAACA
GGGATTTGAGTTCTGTCTGATTCTTGTCTGTCTGCTCGGTCACTTCTTGCGCCGGAAGCA
ACAACAACAACAATGTTGTGTGGTTTTACAGCTCCGTTGGAGTTGGAAGGTTTTG
GTAGTCTGCTAATGGTGGTTTTGTGGGAACAGAGCGAAAGTTCTGAAGCCTTAGAGG
TGTTAGCATCGTCTGGTGCACAGCCTACTCTGTTCCAGAAACGTGCAGCTATGCGTCAGA
GCTCTGGAAGCAAAATGGGAAATTCGGAGAGTTCCGGGAATGAGGAGGTTTAGTGATGATG
GAGATATGGATGAGACTGGGATTGAGGTTTCTGGGTTGAACTATGAGTCTGATGAGATAA
ATGAGAGCGGTAAAGCGGCTGAGAGTGTTCAGATTGGAGGAGGAGGAAAGGTAAGAAGA
AAGGTATGCTGCTAAGAATCTGATGGCTGAGAGGAGAAGGAGGAAGAAGCTTAATGATA
GGCTTTATATGCTTAGATCAGTTGTCCCAAGATCAGCAAAATGGATAGAGCATCAATAC
TTGGAGATGCAATTGATTATCTGAAGGAACCTTACAAAGGATCAATGATCTTCACAATG
AACTTGAGTCAACTCCTCTGGATCTTTGCCTCCAATTCATCAAGCTTCCATCCGTGGA
CACCTACACCGCAAACTCTTTCTTGTGCTGCAAGGAAGAGTTGTGTCCCTCTTCTTTAC
CAAGTCCATAAGGCCAGCAAGCTAGAGTTGAGGTTAGATTAAGGAAGGAAGAGCAGTGA
ACATTATATGTTCTGTGGTCTGACCGGGTCTGTTGCTCGCTACCATGAAAGCTTTGG
ATAATCTTGGATGGATGTTTCAGCAAGCTGTGATCAGCTGTTTTAATGGGTTTGCCTTGG
ATGTTTTCCGCGCTGAGCAATGCCAAGAAGGACAAGAGATACTGCCTGATCAAATCAAAG
CAGTGCTTTTTCGATACAGCAGGGTATGCTGGTATGATCTGATCTGATCCTGACTTCGAGT
CCATTAAGCATCTGTTGAAGCAGAGCTAGAAGAATAAGTCCCTTTAAATCTGCAATTTT
CTTCTCAACTTTTTTTCTTATGTCTAATCTCAATCTAAGCATGTAATGCAATTGCAAAAT

MPPLPSSTAPSSSRHLRSPESIAKFAGRAIFPALQKSCPICLENLTERRSAAVITVCKH
GYCLACIRKWSSFKRNCPLCNTRFDSWFI VSDFA SRKYHKEQLPILRDRETLT YHRNPS
DRRRRIIQRSDRLVENSSTSRPLPWRPSFGRPGSVDPDSVIFQRLKWRASIYTKQLRAVR
LHSRLELSLAVDNIYTKAKITERIEPIWRRELQAVLGDPPDSVIVHFSALFIKRLEREN
NRQTGQGLMVEDEVSLRKFLSDKVDIFWHELRCFAESILTMETYDAVVEYNEVE*

ATGAGTGATTTTGGATGAAAACCTTCATCGAAATGACGTCGTATTGGGCTCCACCATCCAGT
CCTAGCCCAAGAACGATATTGGCAATGCTGGAGCAAACCGACAATGGTCTGAATCCAATC
AGTGAGATCTTCCCTCAAGAAAGCTTGGCCAAGAGATCATACTGATCAATCTGGACAAAGA
TCTGGTCTTCGTGAGAGACTGGCTGCAAGAGTAGGATTCAATCTTCCAACATCAATACA
GAAGAAAACATGAGTCTCTTGGATGCATTTTTCAGGAGCTGCAATGTTTCTTAATCTCC
TCGTGTGCAATCTCTCAGGATTTCAGTCCATCAGCAATATGCATACTCCCAATATGGTC
AGTGATTCTTCCAGATTATCCCTCCGTCTTCAGCCACCAATTACGGACCTCTAGAGATG
GTGGAACCTTCCGGTGAAGACAATGCAGCGATGATGATGTTCAACAACGATCTTCCTTAT
CAGCCGTACAATGTTGATCTGCCCTTCTCTAGAAGTCTTTGATGATATTGCAACGGAAGAG
TCCTTTTATATCCCATCTTATGAACCTCATGTTGACCCAATTGGAACCTCTTTAGTCACA
TCCTTTGAAATCTGAACCTCGTTGACGATGCCCATCCGACATCATCTCATTGAGGACAGT
GAGAGCGAGGATGGAAAACAAGATGATGACGACGAGGACTTCCAATACGAAGACGAAGAC
GAAGACCAATACGACCAAGATCAAGATGTAGATGAAGATGAAGAGGAAGAAAAAGATGAA
GACAATGTTGCATTAGATGATCTTCAACCTCCACCTCCAAGAGAAGGAGATATGAGGTA
TCAAACATGATTGGAGCCACAAGAACAAGCAAGACACAAAGGATCATACTTCAGATGGAA
AGCGACGAAGACAATCCTAACGATGGTTATCGCTGGAGAAAATACGGTCAGAAAGTCGTC
AAAGGAAATCTTAATCCGAGGAGTTACTTCAAGTGCACAAACATCGAGTGCAGAGTGAAA
AAACATGTGAGAGGAGGAGACAGACAATAATCAAGTTGGTTGTGACTACATACGATGGGATA
CACAACCATCTCTTACCACCTGCACGTAGAAGCAATTCAGTTCAAGGAACCGGCTGCA
GGGGCAACAATACCTCAAATCAGAATGATCGAACCAGTCGGTTAGGTAGGGCTCCTCCT
ACTCTACTCTCTACTCTCTCTCTCTCGTCTTACACACCTGAGGAGATGAGGCCCTTC
TCTTCGTTGCTACAGAAATTGATCTGACAGAGGTTTATATGACCCGAATCTCTATGCTG
CCGAATATACCGGTTTACGAGAATTCGGGTTTATGTACCAGAATGATGAACCGACGATG
AATGCGATGCCGATGGTTTCAGATGTGTACGATGGGATCATGGAACGCTGTATTTTAAAG
TTTGGTGTGCACATGTAG

MSDFDENFIEMTSYWAPPSSSPRPTILAMLEQTDNGLNPISIEIFFQESLPRDHTDQSGQR
SGLRERLARVGFNLPTLNTENMSPLDAFFRSSNVPNSPVVAISPGFSPSALLHTPNMV
SDSSQIIPSSATNYGPLEMVETSGEDNAAMMMFNNDLPYQPYNVDLPSLEVFDIATEE
SFYIPSYEPHVDPIGTPLVTSFESELVDDAHTDIISIEDSESEDGNKDDDEDFQYED
EDQYDQDQDVEDEEEKEDNVALDDPQPPPPKRRRYEVSNMIGATRTSKTQRIILQME
SDEDNPNDGYRWRKYGQVVKGNPNPRSYFKCTNIECRVKKHVERGADNIKLVVTTYDGI
HNHPSPPARRNSSSSRNRSAGATIPQNQNDRTSRLGRAPPTPTPTPPSSYTPBEMRPF
SSLAIEDLDEVYMTGISMLPNIPVYENSGFMYQNDEPTMNAMPDGSVDYDGMERLYFK
FGVDM*

>G2555 (177..956)

CTGTTTTTGTATCCGTGTAAATTAATCACACGGTAGTTTTTGTATGAAAAGACAACAATCG
GAGAACAATCTGGTCTGCTGCTAAAAATTAATAAATTGTTTTGTCTAATTGTCTCCACCC
ATAAAAAAGCGCAATTCAATTCACCGACTAAAGACATTCTCCGGTGGAGACCCCGATGC
AATCCACTCATATAAGCGGCGGAAGTAGCGGTGGTGGTGGTGGAGGAGGAGAGGTGA
GTCGAAGTGGATTATCTCGATCCGTTACGCTCCAGCTACTTGGATTGAAACCCCTACTCG
AAGAAGATGAAGAAGAAGGTTTAAACCTAACCTTTGTTAACAGAGCTGCTTACTGGTA
ATAATAACTCTGGAGGAGTGATAACGAGTCGTGACGACTCGTTCGAGTTCCTGAGTTCTG
TTGAGCAAGGATTGTATAATCATCATCAAGGTGGTGGCTTTCACCGTCAGAATAGTTCTC
CGGCTGATTTTCTTAGTGGGTCTGGTTCTGGGACTGATGGGTATTTCTCTAATTTTGGTA
TTCCGGCGAATPATGACTATTTGTGACCAACGTTGATATTTCTCCGACTAAACGGTCTA
GAGATATGGAAACACAGTTTCTTCTCAGCTGAAAGAAGAGCAAATGAGTGGTGGGATAT
CAGGAATGATGGATATGAACATGGACAAGATTTTGGAGATTGAGTTCCTTGTAGGGTTC
GTGCTAAACGTTGGTTGTGCTACTCATCCTCGTAGCATTGCTGAACGGGTGAGAAGAACGC
GAATAAGTGATCGGATTAGGAGGCTGCAAGAGCTTGTTCCTAACATGGATAAGCAAACCA
ACACTGCAGACATGTTGGAAGAAGCTGTGGAGTATGTGAAGGCTCTTCAAAGCCAGATCC
AGGAATTGACAGCAGCAGAGAGATGCAAATGCAAACCTAAAGAAGAACAATAATGTA
TCCTTTAGGATTGATATATCTGTATTTTATTTTGTACTATCTAAAAATGGTGATGATC
TGTTTCGAAAATTCGAAACATGATCTTATATATTGAACTAGAAAAAATAGATATATATGAA
TTTTAGCTGTAAAATTTTGTACAATAAGGAGAAAAAGATTTAGAAGAGTCAATAAAAAAG
ATGATGTTTACAAGTCAAAAAA

>G2555 Amino Acid Sequence (domain in AA coordinates: 175-245)

MQSTHISGGSSGGGGGGEVSRSGLSRIRSAPATWIEITLLEDEEEGLKPNLCLTELLT
GNNNSGGVITSRDDSFELSSVEQGLYNHHQGGGFHRQNSSPADFLSGSGSGTDGYFSNF
GIPANYDYLSNVDISPTKRSRDMETQFSSQLKEEQMSGGISGMMDMNDKIFEDSVPCR
VRAKRGCAHPRISIAERVRRTRISDRIRRLQELVPNMDKQNTADMLEEAVEYVKALQSQ
IQELTEQQKRCCKPKEEQ*

>G375 (53..1171)

TCGACAAAACTCTCACTCTCCCTCAAACATAACACATAACAGAACACAAAATGGGTCT
CACTTCTCTTCAAGTTTGCATGGATTCTGATTGGCTCCAGGAATCCGAGTCATCAGGAGG
AAGCATGTTAGACTCTTCAACGAATTCTCCGTGAGCAGCCGACATACTAGCAGCTTGCAG
CACTAGACCACAAGCCTCGGCCGTGGCTGTAGCCGCTGCAGCTCTGATGGACGGTGAAG
GAGGCTGCGTCCACCTCACGACCATCCTCAAAGTGTCTCGTTGCGAGTCAACACATAC
TAAGTTCTGTACTACAATAACTACAGCCTCTCTCAGCCTCGTTACTTCTGCAAGACTTG
TCGCCGTTACTGGACAAAAGCGGAACCTAAGGAATATTCCGGTTGGTGGTGGATGCCG
TAAAAACAAGAAACCATCTTCTCTAATTCCTCCTCCTCCACTTCTTCCGGCAAAAAACC
ATCCAACATCGTTACCGCCAATACCTCTGATCCTTATGGCTTTAGCACATTCTCATCAAAA
TTACCAACATTCTCCTCTAGGGTTTTCACATTTTGGTGGGATGATGGGGTCTTACTCAAC
TCCGGAGCATGGTAACGTTGGTTTCTTGGAGAGCAAGTATGGCGGTTTGCTTTCGCAGAG
CCCTAGACCTATTGATTTCTTGGACAGTAAGTTTGTATCTCATGGGAGTGAACAATGACAA
CCTGGTCATGGTTAATCATGGAAGTAACGGAGATCATCATCATCATATAATCATCACAT
GGGTCTGAATCACGGTGTAGGTCTTAACAACAACAACAATGGTGGATTTAATGGGAT
TTCTACGGGAGGCAATGGAATGGTGGTGGTCTCATGGATATATCGACATGCCAAAGACT
TATGCTATCTAATTATGATCATCACCATTACAATCATCAAGAAGATCATCAAGGGTAGC
AACAATAATGGATGTGAAGCCTAATCCGAAGTTGTTATCGCTTGATTGGCAGCAAGATCA
ATGCTACTCCAATGGTGGTGGTAGCGGAGGCGCAGGAAAATCCGACGGTGGTGGATACGG
CAATGGTGGTTATATCAACGGTTTAGGTTCTGTCGTGAATGGTTTGTGAATGGCTATGG
AACGTCCACTAAAAEAACTCCTTGGTTTGATAAGTTAATCAGAACTCTTTTTTCTTGT
CGTCATCAACTAGTAGTAGTAGTAATAGTAGTTGGAGACTAGAGAAGCACTTCAAATTAT
TTATGGGTTTGTGCTAAGCCAGTTTAC

>G375 Amino Acid Sequence (domain in AA coordinates: 75-103)

MGLTSLQVCMDSWLQSESSGGSMDSSTNSPSAADILAACSTRPQASAVAVAAAALMD
GGRRLRPPHDPKQKPCRESTHTKFCYNNYSLSQPRYFCKTCRRYWKGGTLRNIPVGG
GCRKNKKPSSNSSSSSTSSGKKPSNIVTANTSDLMALAHSHQNYQHSPLGFSHPGMMGS
YSTPEHGNVGFLESKYGGLLSQSPRPIDFLDSKFDLMGVNNDNLVMVNHGSNGDHHHHHN
HHMGLNHGVLNNNNNNGGFNGISTGGNGNGGGLMDISTCQRLMLSNDHHHHYHQEDHQ
RVATIMDVKNPKLLSLDWQDDQCYNNGGSGGAGKSDGGGYNGGYYINGLSSWNGLMN

GYGTSTKTNSLV*

>G1007 (86..763)

ATTCCTTCTTGCCCTAGGAACATAATTGTTGCACACTTCGGTACACAATTTTTTGAGCACTT
CGACATCAAAACGAGAGAGAAAAGAATGGTGGATTCTCATGGCTCCGACACGGAATGTTT
CTCCAAGAAGAAAAAGGAGAAAAACGAAAGAAAAGGGGTATATCGTGGGGCTCGCATGAG
GAGCTGGGGGAAATGGGTCTCGGAGATTGGGGAGCCCCGTAAGAAATCAAGAATCTGGCT
CGGGACTTTCCCAACGGCGGAGATGGCAGCGCGTGCCCATGATGTTGCGGCATTGAGTAT
CAAAGGAAGTTCGCAATCCTTAACCTCCCTGAGCTCGCGGATTTTCTGCCAAGACCAGT
CTCGCTCAGCCAACAGGATATCCAGGCCGAGCCGCCGAAGCCGCTCTTATGGATTTCAA
AACTGTACCATTCATCTTCAGGATGACTCAACGCCGTTGCAAACTAGGTGTGATACTGA
GAAGATCGAAAAGTGGTCATCCTCATCGTCTCAGCCTCATCCTCATCCTCATCTTCGTC
CTCGTCTCATCATCTATGCTTTCGGGGGAGCTAGGAGATATTGTGGAGTTGCCGAGTCT
TGAAAACAATGTAAATACGATTGTGCGCTGTATGACTCGTTGGAGGGGCTGGTGTGAT
GCCCCCATGGTTAGATGCTACCGAAAATGATTTTAGGTATGGAGATGATTCGGTACTGTT
GGACCATGTCTCAAAGAAAGCTTTTTGTGAATTATGAGTAAGGTTTTTTTTTGGAAAG
AAATGTGGTTTTTTGTTTCTCTCTCTTTTATACTTTTCGATCTTTTTTTCTAAGCATAT
ATATCTTCTACATATGTAATACTTTTCCATTAGTAAACAATGATTCGGTTTCGGGTACAA
AAAAAAAAAAAAAAAAAAAAA

>G1007 Amino Acid Sequence (domain in AA coordinates: TBD)

MVDSHGSDTECSSKKKKEKTKKEGVYRGARMRSWGKQVSEIREPRKKSRIWLGTFPTAEM
AARAHDAALSIKSSAILNFPBLADFLPRPVSLSQDIIQAAAAEALMDFKTVPFHLQD
DSTPLQTRCDTEKIEKWSSSSSSSSSSSSSSSSSSSSSMLSGELGDIVELPSLENNVKYDC
ALYDSLEGLVSMPPWLDATEFRYGDSDVLLDPLCKESFLWNYE*

>G1010 (344..1276)

ATTCCTTCTTCTAAAAAATCTTGACAACTTTTGTTTTGTTTTCTTTCTCTGAATTTTTT
AAAAGAGAGAGAGCTATGTAGCTATGAAACAGTAAGAGATATAGATATAGAGAGACAGAG
AAAGATGATGATCAGTGAAGTTAGGCTAAACCCACTTTCTATTTATGTATAATTAGGTCA
ATCACATACCAATCTCCTCCTCAATTCTCCTCCTCCTCCTCCTCAAAATCTAGGGTTTTG
CTTGATCTCACCCCTTTCTCAATCCCTAGGGAACTGTGAATTTTCATCAAAATCCAT
TATTTTTTGGTCACACCTTAAAGAGATCTGAGAGTTCTAAAGATGATGACAGATTTATC
TCTCAGGAGAGATGAAGATGAAGAAGCAAGCCCTTAGCAGAAGAAGAGAGCGCG
TGAAGTAGCAGACAGAGAGACATGTTTCGACAAAGTTGTGACTCCAAGTGATGTCGGAAA
ACTAAACCGACTTGTGATCCCAAAGCAACACGAGAGAGATTCTTCCCTTTAGATTCATC
TTCAAACGAGAAAGGTTTGTCTTTTAACTTCGAAGATCTCACTGGCAAATCTTGGAGGTT
CCGTTACTCTTACTGGAACAGTAGTCAAAGCTATGTCATGACTAAAGTTGGAGCAGATT
CGTTAAAGACAAAAAGCTTGACGCCGAGATATTGTCTCTTTCCAAAGATGTGTGCGAGA
TTCAGGAAGAGATAGCCGTTTGTATTATTGATTGGAGGAGAAGACCTAAAGTCCCTGACCA
TCCTCATTTCCGCCCGGAGCTATGTTCCCTAGGTTTTACAGCTTTCCTTCGACCAATTA
CAGTCTTTATAATCATCAGCAGCAACGTCATCATCAGTGTTGGTGGTTATAATTATCA
TCAAATTTCCGAGAGAAATTTGGTTATGGTTACTTCGTTAGGTCACTGGATCAGAGGAACAA
TCCTGCGGCTGCGGTGGCTGATCCGTTGGTGATTGAATCTGTGCCGGTGATGATGCACGG
GAGAGCTAATCAGGAACCTGTTGGAACGGCCGGAAGAGACTGAGGCTTTTTTGGAGTTGA
TATGGAATGCGGCGAGAGCGGAATGACCAACAGTACGAGGAGGAATCATCATCTTCCGG
TGGAAGTTTGCCACGTGGAGGCGGTGGTGGTCTTCATCTTCTCTTTTCAGCTGAG
ACTTGAAGCAGCAGTGAAGATGATCACTTCACTAAGAAAGGAAAGTCTTCATTGTCTTT
TGATTGGATCAATAATAATGATGATGATGAAATTAGTTGGTATTTTAAGAAAAAACA
TACATATATAATTCTATATATATGACAACATAATGCATTGATTTCCCTT

>G1010 Amino Acid Sequence (domain in AA coordinates: 33-122)

MMTDLSLTRDEDEEEAKPLAEEEGAREVADREHMFQVVTTPSDVGKLNRLVIPKQHAERF
FPLDSSSNEKGLLLNFEDLTGKSWRFYSYWNSSQSYVMTKGWSRFVKDKKLDAGDIVSF
QRCVGDSDGRDRLFIDWRRRPKVPDHPHFAAGAMFPRFYSFPSTNYSLYNHQQQRHHHS
GGYNYHQIPREFGYGYFVRSVDQRNPNAAVADPLVIESVPVMMHGRANQELVGTAGKRL
RLFGVDMCEGESGMTNSTBEESSSSGSLPRGGGGGASSSSFFQLRLGSSSEDDHFTKKG
KSSLSFDLDQ*

>G1014 (174..1112)

CACAAACCACAGTCTCTCTTCTCTCTATCTATCTTCTCTTCTCTCTCTATCTCTAT
CACTGAAACCCAAAGAGATCCACCATTTGTTCTTTTTTCTCTTACACAGAGAACTGTTT

CTTCCACACTTCCTTTTACTAGGCAGTGTTAACCAATTGAGAGAGAAAAATGATGGTTG
ATGAAAATGTGGAAACCAAGGCCTCTACTTTAGTGGCAAGTGTGATCATGGGTTTGGAT
CCGGGTCGGGTCATGATCATCATGGGTTATCGGCGTCTGTGCCTCTTCTTGGTGTAACT
GGAAGAAGAGAAGGATGCCTAGACAGAGACGATCTTCTTCTTCTTTAACCTTCTCTCTT
TCCCTCCTCTATGCTCCTATTTCCACGTGCCAATCCTCTCCCGCACGTAAAATTG
ACCCAAGAAAGCTAAGATTCTCTTCCAAAAGGAAGTCAAGAACAGTGACGTGAGCTCTC
TCCGACGTATGATACTCCCGAAGAAAGCCGCGGAGGCTCACTTGCCGGCAGTTGAATGCA
AGGAAGGATTCCCTATAAGAAATGGAAGATTGGACGGTTTTACGTTTTGGACCTTCAAGT
ATAGGTACTGGCCAAACAACAATAGCAGAATGTACGTGCTAGAAAACACAGGCGATTTTG
TGAATGCTCATGGTCTGCAGCTAGGTGACTTCATCATGGTTTACCAAGATCTCTACTCAA
ACAATTACGTTATACAAGCAAGAAAAGCATCGGAAGAAGAAGTAGACGTAATCAATC
TTGAAGAAGACGACGTTTACACAAACTTAACAAGGATCGAAAACACTGTGGTTAACGATC
TTCTCCTCCAAGATTTTAATCATCACAACAACAACAACAACAGCAACAGCAACA
GCAACAAATGTTCTTACTATTATCCAGTCATAGATGATGTCACCACAAACACAGAGTCTT
TTGTCTACGACACGACGGCTCTTACCTCCAACGATACTCCTCTCGATTTTTTGGGTGGAC
ATACGACGACTACTAATAATTATTACTCCAAGTTCGGAACATTGATGGTTTGGGCTCCG
TTGAGAATATCTCTCTCGATGACTTCTACTAGATAATCAATCGATGGGCTCATGGTATTC
TTGATGGTGATCAGCTATTTAATATCCTTATAATATATATAAGAATTAAATGCAATTTGC
ATATATATTATCAAGTGTGTGAATATAACATTACAGTTTTAAAAAAAAAAAAAAAAAAAA
>G1014 Amino Acid Sequence (domain in AA coordinates: 90-172)

MVDENVETKASTLVASVDHGFSGSGHDHGLSASVPLLGVNWKRRMPQRSSSSFNL
LSFPPMPPIISHVPTPLPARKIDPRKLRFLFQKELKNSDVSLRRMILPKKAAEAHLPAL
ECKEGIPIRMEDLDGFHVWTFKYRYWPNNNSRMYVLENTGDFVNAHGLQLGDFIMVYQDL
YSNNYVIQARKASEEEVDVINLEEDDVYTNLRIENTVNDLLQDFNHNNNNNNNSN
SNSNKCSYYPVIDDVTTNTESFVYDITALTSTNDTPLDFLGGHTTTTNNYYSKFGTFDGL
GSVENISLDDFY*

>G1035 (103..624)

CCATAATAATATATAAACTATATACTATAATCTTTTTACATAATAAACTTTGGGTCTT
GCGTCTTAATCATAGTACTTAATTTCTCTGTGTGTTTTAATATGAATAATAAACTGAA
ATGGGATCTTCCACAAGTGGAATTGCTCGTCGGTTTTCAACCACTGGTTTAGCTAACTCC
GGTTCAGAATCTGATCTCCGGCAACGTGATCTAATCGACGAGCGGAAGAGAAAGAGGAAA
CAGTCGAACAGAGAATCTGCGAGGAGGTCGAGGATGAGGAAGCAGAAGCATTTGGATGAT
CTCACTGCTCAGGTGACTCATCTACGTAAAGAAAACGCTCAGATCGTCGCGGAATCGCC
GTCACGACGCAGCACTACGTCACTATCGAGGCGGAGAACGACATTCTCAGAGCTCAGGTT
CTTGAACCTTAACCACCGTCTCCAATCTCTTAACGAGATCGTTGATTTCTGCGAATCTTCT
TCTTCAGGATTCCGTATGGAGACCGGTGAGGATTATTTCGCGGTGGATTATTCGACGGC
GTGATGAATCCTATGAATCTAGGGTTTTATAATCAACCAATCATGGCTTCTGCTTCTACT
GCTGGTGATGTTTTCAACTGTTAGAAAACCTCACATCATTATCATCGTGAGTGAGACTAA
TCATCGCAGCAGGGGTAAAACGTAAATTTTTCTTATAAATTATGTGATGATGCTTTGTTT
CTTTATTTTATAAGATGGTTAATTAGTGTTTAAAACGATTGTAATGATAGACAGTGTA
GAAATGTGTGATATCATGGAGATGGTGATGTGAGTTTGGTACAAATATTTAAGATCTTT
TCTTTCTATATATTAAGTGAAGAAATAATATTTGTCAATTTCTTAAAAAAAAAAAAAA
AAA

>G1035 Amino Acid Sequence (domain in AA coordinates: 39-91)
MNNKTEMSSSTSGNCSSVSTTGLANSSESIDLQRDLIDERKRKRKQSNRESARRSRMRK
QKHLDDLTAQVTHLRKENAQIVAGIAVTTQHYVTIEAENDILRAQVLELNHRLQSLNEIV
DFVESSSSSGFMETGQGLFDGGLFDGVMNPMNLGFYNQPIMASASTAGDVFNC*

>G1046 (1..567)

ATGATTAGACATCTAAAACCTACATGGAGTCGTCTAGTGTCATCGCTCTCATTTGTTTC
GATATTCTTGATGGAGTCCCACTACACGACGATCATTTCAACTCGGCATTCTACCAAAC
ACTGACTTTAATGTTTCAATTTGCAGTCAAACGTATCGACCCGCATCAACAATCAGTCTCAC
TTAGACCCAAATGCAGAAAACATTTTCCATAACGAAGGTCTTGCTCCAGAAGAAAGAAGA
GCAAGAAGAATGGTCTCTAACCGGGAATCTGCAAGGAGGTACGTATGCGCAAAAAGAAG
CAGATCGAAGAGCTGCAACAACAAGTTGAACAACCTCATGATGTTGAATCATCACTTGTCT
GAGAAAGTCATCAACTTGTGGAAAGCAACCATCAGATCCTACAAGAGAACTCACAGCTG
AAAGAGAAAGTCTCTTCTTCACTTGCTCATGGCAGATGTGCTATTACCCATGAGAAAT
GCAGAGAGCAACATCAATGACCGCAATGTGAATTATCTAAGAGGAGAACCATCAACCGT

CCCACCAACAGTCCCTTTGGTAAGTAA

>G1046 Amino Acid Sequence (conserved domain in AA coordinates:79-138)

MIRHLKPYMESSSVHRSHCFDILDGVPLHDDHFNSAFLPNTDFNVHLQSNVSTRINNQSH
LDPNAENIFHNEGLAPEERRARMVSNRESARRSRMRKKQIBELQQQVEQLMMLNHHLS
EKVINLLESNHQILQENSQLKEKVSSFHLLMADVLLPMRNAESNINDRNVNVLRGEPNSNR
PTNSPFGK*

>G1049 (29..550)

CTAACTTTCTTCCCAAGTAAACTTCAAAATGCAGCCGCAAACAGACGTTTTCAGCCTCCA
TAACTACCTAAACTCATCGATACTGCAGTCTCCGTATCCTTCTAATTTCCCGATATCTAC
GCCATTTCCAACCAACGGTCAAAACCCGTACCTCCTCTACGGATTCCAAGCCCTACAAA
CAATCCACAATCCATGAGCCTAAGCAGCAACAACCTCAACATCAGATGAAGCAGAAGAGCA
GCAGACGAACAACAAATAATCAACGAGCGGAAGCAGAGAAGGATGATTTCAAACCGAGA
ATCCGCAAGGAGATCGCGTATGAGGAAGCAAAGACACCTTGACGAGCTTTGGTCACAAGT
GATGTGGTTAAGGATCGAGAATCATCAGTTGCTTGATAAGCTTAACAATCTCTGAGTC
TCACGACAAGGTTCTTCAAGAGAATGCTCAGCTTAAAGAAGAAACATTTGAGCTTAAGCA
AGTGATCAGCGATATGCAAATCAAAGCCCTTTCTCTTGCTTTAGAGACGATATAATCCC
CATTGAATAAAGCATTTTCCCGGATTCAATTTATGAAAATTTCTTCAAGAGTATGTT
TCTTTGTATGTATATGTGGAGATGTATTTCAAGGTTTGTATAATATGACCCTTACGACG
ACGTTTTTAGATTGTAGTAAATTTATAAACTAAAGAAGATTAGTGTTAATGAAGAACAAA
TATAA

>G1049 Amino Acid Sequence (domain in AA coordinates 77-132)

MQPQTDVFLSHNYLNSILQSPYPNFIPTFPPTNGQNPYLLYGFQSPNTNPQSMSSLSS
NNSTSDEAEEQQTNNNIINERKQRRMISNRESARRSRMRKQRLHDELWSQVMWLRINHQ
LLDKLNNLSESHDKVLQENALKEETFELKQVISDMQIQSPFSCFRDDIPIE*

>G1069 (89..934)

TTGGAACCTTAGAGGCTTTCAAGCAAATCATCAGGGTAACAATTTCTTGATCTTTCTTT
TTAGCGAATTTCCAGTTTGTGTCATCATGGCAAACCTTGGTGGACGAACCAGAGTGG
TTTAGCGGCATGGTGGACCATTCGGTCTCCTCAGGCCATCACCAAAACCATCACCACCA
AAGTCTTCTTACCAAAGGAGATCTTGAATAGCCATGAATCAGAGCCAAGACAACGACCA
AGACGAAGAAGATGATCTTAGAGAAGGAGCCGTTGAGGTGGTCAACCGTAGACCAAGAGG
TAGACCACCAAGGATCCAAAAACAAACCCAAAGCTCCAATCTTTGTGACAAGAGACAGCCC
CAACGCACTCCGTAGCCATGTCTTGAGATCTCCGACGGCAGTGACGTCGCCGACACAAT
CGCTCACTTCTCAAGACGCAAGGCAACGCGCGTTTGCCTTCTCAGCGGGACAGGCTCAGT
CGCTAACGTCACCCTCCGCCAAGCCGCGCACCAGGAGGTGTGGTCTCTCTCAAGGCAG
GTTTGAAATCTTATCTTTAACCGGTGCTTTCTCCTCCCTGGACCTTCCCCACCCGGGTCAAC
CGGTTTAAACGGTTTACTTAGCCGGGGTCCAGGGTCAGGTGCGTTGGAGGTAGCGTTGTAGG
CCCACTCTTAGCCATAGGGTCCGGTCATGGTGATTGCTGCTACTTTCTCTAACGCTACTTA
TGAGAGATTGCCCATGGAAGAAGAGGAAGACGGTGGCGGCTCAAGACAGATTACCGGAGG
CGGTGACTCACCGCCAGAATCGGTAGTAACCTGCCTGATCTATCAGGATGGCCGGGCC
AGGCTACAATATGCGCGCCGATCTGATTCCAAATGGGGCTGGTCAGCTAGGGCACGAACC
ATATACATGGGTCCACGCAAGACCACCTTACTGACTCAGTGAGCCATTTCTATATATAAT
GGTCTATATAAATAAATATATAGATGAATATAAGCAAGCAATTTGAGGTAGTCTATTACA
AAGCTTTTGTCTGCTGGTTGGAAAAATAAATAAGTATCAAAGCTTTGTTTGTCTTAAATGGA
AATATAGAGCTTGGGAAGGTAGAAAGAGACGACATT

>G1069 Amino Acid Sequence (domain in AA coordinates: 67-74)

MANPWWTNQSLAGMVDHSVSSGHHQNHQSLLTGDLGIAMNQSQDNDQDEEDDPREG
AVEVNNRRPRGRPPGSKNPKAPIFVTRDSPNALRSHVLEISDGSVDVADTIAHFSRRRQR
GVCVLSGTGSVANVTLRQAAPGGVVSLQGRFEILSLTGAFPLPGSPPGSTGLTVYLAGV
QQQVVGGSVVGPLLAIGSVMVIAATFSNATYERLPMEEEDGGGSRQIHGGDSPPRIGS
NLPDLSGMAGPGYNMPPHLIPNGAGQLGHEPYTWVHARPPY*

>G1070 (170..1144)

TCGACCAGCTTGGATTTCTGTTGTCATCTACTACTCTCTTTCTTCTCTAGCTAGCTA
GTTTTGACAGCAAAATAAGAAAGCAAAAAAGGTCAACTAAAAAAGATCTGTTCTTAGAT
CACTCTCTTCTTTCTTTTGGATCCAAATCCACCATTGAATCATAGATCATGGATCCAGT
ACAAATCTCATGGATCACAAAGCTCTCTACCTCCTCCTTTCCACGCAAGAGACTTTCAATT
ACATCTTCAACAACAGCAACAAGAGTTCTTCTCCTCCACCATCACCAGCAACAAAGAAACCA
AACCAGTGGTGACCAACAAGGAGGATCAGGAGGAAACCGACAAATCAAGATGGATCGTGA

AGAGACAAGCGACAACATAGACAACATAGCTAACAAACAGCGGTAGTGAAGGTAAAGACAT
AGATATACACGGTGGTTTCAGGAGAAGGAGGTGGTGGCTCCGGAGGAGATCATCAGATGAC
AAGAAGACCAAGAGGAAGACCAGCGGGATCCAAGAACAACCAAACCACCGATTATCAT
CACACGGGACAGCGCAAAACGCGCTTAGAACCCACGTGATGGAGATCGGAGATGGCTGCGA
CTTAGTCGAAAGCGTTGCCACTTTTGCACGAAGACGCCAACGCGCGTTTGCCTTATGAG
CGGTACTGGAATGTTACTAACGTCATATACGTACGCTGGATCTCATCCTTCTCCTGG
CTCGGTAGTTAGTCTTACCGGAAGGTTTCGAGATTCTATCTCTCTCAGGATCTTTTCTCCC
TCCTCCGGCTCCTCTACAGCCACCGGATTGAGTGTTTACCTCGCTGGAGGACAAGGACA
GGTGGTTGGAGGAAGCGTAGTTGGTCCGTTGTTATGTGCTGGTCTGTGCTTGTATGGC
TGCGTCTTTTAGCAATGCGGCGTACGAAAGGTTGCCTTTAGAGGAAGATGAGATGCAGAC
GCCGGTTCATGGCGGAGGAGGAGGAGGATCATTGGAGTCCGCCCAATGATGGGACAACA
ACTGCAACATCAGCAACAAGCTATGTCAGGTCATCAAGGGTTACCACCTAATCTTCTTGG
TTCGGTTAGTTGACGACGAACATGATCAGTCTTATTTGGTCAACGGGACGACCACCGTA
TTGATCAATATACACACACTCATAATCGTTGCTAGCTAGCTAACGATGAATCATGAG
TTTAGTGGATATATATATGATTAAGAGGTTAGCTTATGAACATTAATAAGAGTTTGA
TTCTATCGAGCTTCATTATGTTTGGGTCATCGTTC

>G1070 Amino Acid Sequence (domain in AA coordinates: 98-120)
MDPVQSHGSQSSLPFPHARDPQLHLQQQQQEFFLHHHQQQRNQTGDQGGSGGNRQIK
MDREETSNDINNIANNSSSEKDDIDIHGSGSEGGGSGGDHQMTRRPRGRPAGSKNPKPK
PIIITRDSANALRTHVMEIGDGC DLVESVATFARRRQRGVCMVSGTGNVTNVTIRQPGSH
PSPGVS VSLHGRFEILSLSGSFLPPPAPPTATGLSVYLAGGQGVVGGSVVGPLLCAGPV
VMAASF SNAAYERLPLEEDEMOTPVHGGGGGSGLESPPMMGQQLQHQQQAMSGHQGLPP
NLLGSVQLQQQHDQSYWSTGRPPY*

>G1076 (198..1076)
ATTTTAGTCTTCTATAACTTCTTCTCAATCCTCTCTCATATCTTTTTTCTTAGTTTAAA
TTTCAATAAAATAGAAAAAACATATACAAATCTACAGAGAAGAGAAGCTTTATTTTAAT
CTTGTGTGTGTGTGTGTGTTTTATATAATTTTTATTTTTCCTTCAATTAATAATCTCTCT
TTGCTTTTGATGTGGGCATGGCTGGTCTTGATCTAGGCACAGCTTTTCGTTACGTTAATC
ACCAGCTCCATCGTCCGATCTCCACCTTCACCACAATTCCTCTCCGATGACGTCACCTC
CCGGAGCCGGGATGGGTCAATTTACCGTTCGACGACGAAGACAACAACAACCATCAAG
GTCTTGACTTAGCCTCTGGTGGAGGATCAGGAAGCTCTGGAGGAGGAGGATCACGGCG
GGGGAGGAGACGTCGTTGGTTCGTCGTCACGTCGACGACGACGACGACGACGACGACGAC
CGAAACCTCCGGTAATTATCAGCGCGAGAGCGCAAACTCTAAGAGCTCACATTCTTG
AAGTAACAAACGGCTGCGATGTTTTGACTGCGTTGCGACTTATGCTCGTCGGAGACAGC
GAGGGATCTGCGTTCTGAGCGGTAGCGGAACGGTCACGAACGTCAGCATACGTCAGCCAT
CTGCGGCTGGAGCGGTTGTGACGCTACAAGGAACGTTCCGAGATTCTTCTCTCTCCGGAT
CGTTTCTTCTCTCTCCGGCACCTCCCGGAGCAACGAGTTTGACAATTTTCTTAGCCGGAG
GACAAGGTCAGGTGGTTGGAGGAAGCGTTGTGGGTGAGCTTACGGCGGCTGGACCGGTGA
TTGTGATTGCAGCTTCGTTTACTAATGTTGCTTATGAGAGACTTCCTTTAGAAGAAGATG
AGCAGCAGCAACAGCTTGGAGGAGGATCTAACGGCGGAGGTAATTTGTTTCCGGAGGTGG
CAGCTGGAGGAGGAGGAGGACTTCCGTTCTTTAATTTACCGATGAATATGCAACCAAATG
TGCAACTTCCGGTGGAAGGTTGGCCGGGGAATTCGGGTGGAAGAGGTCTTTCTGATGTG
TATATATTGATAATCATTATATATATACCGCGGAGAAGCTTTTCCGGCGAAGAATTGTC
GAGAGTGAAGAAAGGTTAGAAAAGCTTTTAATGGACTAATGAATTTCAAATTATCATCGT
GATTTCCGGACATTGTCTTGTTCATCATGTTAAGCTTAGGTTTATTTTGTGCTTGTAG
AATTTTATGTTGAATCTTTTTTTTTTCTGTGAACTCTATTGTGTTCTGCTGCGAAGG
AAAAAAAATTCTCAAAAAAA

>G1076 Amino Acid Sequence (domain in AA coordinates: 82-89)
MAGLDLGTAFRYVNHQLHRPDLHLHNSSSDDVTPGAGMGHFTVDEDDNNNNHQLDLAS
GGGSGSSGGGGHGGGDDVVGRRPRGRPPGSKNPKPPVITRESANTLRAHILEVTNGC
DVFDCAVYARRRQRGICVLSGSGTVNVSIRQPSAAGAVVTLQGTFFELSLSGSFLPPP
APPGATSLTIFLAGGQGVVGGSVVVELTAAGPVIVIAASFNTVAYERLPLEEDEQQQQL
GGGSNNGGNLFP EVAAGGGGLPFFNLPMNMQPNVQLPVEGWPGNSGGRGPF*

>G1089 (31..2427)
AAGTAAGAGAGCTTCTTAAGGAAGAAGAAGATGGGTTGTGCTCAATCAAAGATCGAGAAC
GAAGAAGCAGTTACTCGTTGCAAAGAACGAAAACAATTGATGAAAGACGCCGTCCTGCT
CGTAACGCTTTCGCCGCCGCTCACTCAGCTTACGCTATGGCTCTTAAAAACACCGGAGCT

GCTCTTTCCGATTACTCTCACGGCGAGTTTTTAGTCTCTAATCACTCGTCTTCCTCCGCA
GCTGCAGCAATCGCTTCTACTTCTTCTTCCCCTGCTATATCTCCTCCTTCTTCTCT
TCCACCGCTCCGGTTTCTAATTCACCGCTTCTTCTTCTCCTCCGCTGCGGTTCCCTCAGCCG
ATTCTTGATACTCTTCTCCTCCTCCTCCTCCTCCACCACCGCTTCTTCTTCAACGTGCTGCT
ACTATGCCGAGATGAACGGTAGATCCGGTGGTGGTCATGCTGGTAGTGGACTCAACGGA
ATTGAAGAAGATGGAGCCCTAGATAACGATGATGATGACGATGATGATGATGATGACTCT
GAAATGGAGAATCGTGATCGTTTGATTAGGAAATCGAGAAGCCGTGGAGGTAGTACTAGA
GGAAATAGGACGACGATTGAAGATCATCATCTTCAGGAGGAGAAAGCTCCGCCACCTCCC
CCTTTGGCGAATTCGCGGCCAATTCCGCCGCCACGTGAGCATCAGCATCAACATCAGCAA
CAGCAACAACAACCTTTCTACGATTACTTCTTCCCTAATGTTGAGAATATGCCTGGAACT
ACTTTAGAAGATGATCTCTCACAACCACAACCACAACCACAAGGCCTGTGCCTCCTCAA
CCACATTACCAGTCGTTACTGAGGATGACGAAGATGAGGAGGAGGAAGAGGAGGAAGAG
GAGGAGGAAGAGGAGACGGTGATTGAACGGAACCACTGGTGGAGGAAAGACCGAAGAGA
GTAGAGGAAGTGACGATTGAATTGGAAAAAGTTACTAATTTGAGAGGGATGAAGAAGAGT
AAAGGGATAGGATTCCCGGAGAGAGGAGAGGAATGCGAATGCCGGTGACTGCGACGCAT
TTGGCGAATGTATTGAGCTTGATGATAATTTCTTGAAAGCTTCTGAAAGTGCTCAT
GATGTTTCTAAGATGCTTGAAGCTACTAGGCTCCATTACCATTCTAATTTGCAGATAAC
CGAGGACATATTGATCCTCTGCTAGAGTGATGCGTGTAATTACATGGAATAGATCATTT
AGAGGAATACCAATGCTGATGATGGGAAAGATGATGTTGATTGGAAGAGAATGAACT
CATGCTACTGTTCTTGACAAATTGCTAGCATGGGAAAAGAAGCTCTATGACGAAGTCAAG
GCTGGCGAATCATGAAAATCGAGTACCAGAAAAGGTTGCTCATTTAAATCGGGTGAAG
AAACGAGGTGGCCACTCGGATTCATTAGAGAGAGCTAAAGCAGCAGTAAGTCATTTGCAT
ACAAGATATATAGTTGATATGCAATCCATGGACTCCACAGTTTCAGAAATCAATCGTCTT
AGGGATGAACAATACTAAAGCTCGTTACCTTGTGAGGCGATGGGGAAGATGTGG
GAAATGATGACAAATACATCATCAAAGACAAGCTGAGATCTCAAAGGTTGAGATCTCTA
GATGTTTCAACAAGCGGTGAAAGAAAACAATGATCATCATCACGAACGCACCATCCAGCTC
TTGGCAGTGTTTCAAGAATGGCACACGCAGTTTTCAGGATGATAGATCATCAGAAAGAA
TACATAAAGCACTTGGCGGATGGCTAAAGCTAAATCTCATCCCTATCGAAAGCACACTC
AAGGAGAAAGTATCTTCGCCTCCTCGAGTTCCCAATCCCGCAATCCAAAACTCCTCCAC
GCTTGGTATGACCGTTTAGACAAAATCCCCGACGAAATGGCTAAAAGTGCCATAATCAAT
TTCGACGCGGTTGTAAGCAGGATAATGCAGCAGCAAGAAGACGAGATAAGTCTCAGAAAC
AAATGCGAAGAGACAGAAAAGAAATTGGGAAGAAAAATTAGACAGTTTGAGGATTGGTAC
CACAAATACATCCAGAAGAGAGGACCGGAGGGGATGAATCCGGATGAAGCGGATAACGAT
CATAATGATGAGGTCGCTGTGAGGCAATTCAATGTAGAACAATTAAGAAGAGGTTGGAA
GAAGAAGAAGAAGCTTACCATAGACAAAGCCATCAAGTTAGAGAGAAGTCACTGGCTAGT
CTTCGAACTCGCCTCCCGGAGCTTTTTCAGGCAATGTCCGAGGTTGCGTATTTCATGTTG
GATATGTATAGAGCTATAACGTATGCGAGTAAGCGGCAAAGCCAAAGCGAACGGCATCAG
AAACCTAGCCAGGACAGAGTTTCGTAAGAACAATGTAAAGATCAGAGTAATGTCTTCTC
TTCTTTGATCTTTGAATTTTAAAGCACACACATACATACAACGTATAGCTAAATCTTTATC
ATTGCTTTCTTATATTAAGGTTTGGCTTTTGTAAAGAGGTTTCTTACATATGAGATTCA
TATAGTGTGTTGATTCTTAAGGAAGTCTGTTGAGTAATAAGAAAGTTGTGTATTGAAA
TAGAGTTGCATTTGTTAATTTTG

>G1089 Amino Acid Sequence (domain in AA coordinates 425-500)

MGCAQSKIENEEAVTRCKERKQLMKDAVTARNAFAAAHSAYAMALKNTGAALSDYSHGEF
LVSNHSSSSAAAAIASTSSLPTAISPLPSSTAPVSNSTASSSSAAVPQPIPDLPPLPPPP
PPPLPLQRAATMPPEMNGRSGGGHAGSGLNGIEEDGALDNDDDDDDDDDSEMNDRDLIR
KRSRGGSTRGNRTTIEDHHLQEEKAPPPPLANSRPIPPRQHQQHQQQQQQQFFDYDF
FPNVENMPGTTLEDFFPQPQPQPTRPVPPQPHSPVVTEDDEDEEEEEEEEEEEETVIER
KPLVEERPKRVEEVTEIELEKVTLNLRGMKSKGIGIPGERRGMRMPVTATHLANVFIELDD
NFLKASESAHDVSKMLEATRLHYHSNFADNRGHIHDSARVMRVITWNRFRGIPNADDGK
DDVDLEENETHATVLDKLLAWKKLYDEVKAGELMKIEYQKKVAHLNRVKKRGHSDSLE
RAKAASVSHLHTRYIVDMQSMDSSTVSEINRLRDEQLYLKLVHLVEAMGKMWEMMQIHHQRQ
AEISKVLRLSDVSQAVKETNDHHHERTIQLLAVVQEWHTQFCRMIDHQKEYIKALGGWLK
LNLPIESTLKEKVSSPPRPVNPAPAIQKLLHAWYDRLDKIPDEMAKSAIINFAAVVSTIMQ
QQEDEISLRNKCEETRKELEGRKIRQFEDWYHKYIQKRGPEGMNPDEADNDHNDEVAVRQF
NVEQIKKRLEEEEEAYHRQSHQVREKSLASLRTRLPQLFQAMSEVAYSCSDMYRAITYAS
KRQSQSERHQKPSQGGSS*

>G1093 (1..531)

ATGGGTTATCCGGTGGGTACACTGAGCTCCTCCTCCCAAGAATCTTCCTTCACTTACTC
TCTCTCTTAGGCTTAATACGAACACTCATAGACACGGGTTTTTCGGATATTGGGTCTACCC
GACTTTCTCGAATCCGACCCGGTTTCATCGTCATCGTCATGGCTGGAACACCGTATATG
TCCACGGCGGCGCATCATACCAAGAAAGCTCATTTTTCTTCCCAGTGGCGGCGAGGCTA
GCTGGAGAAATCTTGCCCGTCATCAGATTCTCGGAGCTAACTCGACCCGGATTCCGATCC
GGATCCGATTGCTGCGCGGTGTGCCTCCACGAGTTCGAGAACGATGACGAGATCCGACGG
CTGACGAATTGTCAACACATATTTACCCGGAGCTGTTTAGACCGTTGGATGATGGGTTAT
AATCAGATGACGTGTCCACTTTGTAGAACGCCGTTTATTTCTGATGAGTTACAAGTTGCT
TTTAACCAACGAGTTTGGTCTGAATCTGAACCTTCTCGCAGAATCAAATTAG

>G1093 Amino Acid Sequence (domain in AA coordinates: 105-148)

MGYPVGYTELLLPRIFLHLLSLGLIRTLIDTGFRILGLPDFLESPPVSSSSSWLEPPYM
STAAHHHQESSFFFPVAARLAGEILPVIRFSELTRPGFGSGSDCCAVCLHEFENDDEIRR
LTNCQHI FHRSCLD RWMGMYNQMTCP LCRTPFISDELQVAFNQRVWSESELLAESN*

>G1127 (191..1351)

GACAGACTCTCTGTATGTGTGCGAGAAGCGAGAAGCGAGAGAGAGAGAGAGAGAGTTG
TTAGCTCAGCGCTTTCTCTATTTTCTCGGAATTCACAAAACAGAAAGTTTCATCCTTTA
CGAGAATTAAAGCCGAAAGAAACAATCTTTGAGTTTGATTCTTCTTCTTCTTCTTCTCT
CTCTGCTCTAATGGATTCCAGAGACATCCACCGTCACATAACCAGCTTCAACCACCACC
GGGAATGTTAATGTCTCATTACCGTAACCTAACGCCGCGCTTACCATTAAATGGTTCC
CACTTCCACATCTCAACCGATTCAACACCCTCGTCTTCTTTTGGCAATCAACAACAATC
TCAAACGTTTCATCAGCAGCAACAACAACAATGGATCAGAAGACTCTTGAATCTCTTGG
ATTTGGTGATGGATCACCCTTCTTCTCAACCGATGCGATTCCGGATCGATGATCAGAATCA
GCAACTGCAAGTGAAGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
TGCTTTAGGTTTAGCTCCTACGTCTCTCTTCTCTCTGACGTTCTAATTCTTACGGTGA
GGGTGGTGTGTGAGATAGTGGTGGAAATGGAACTCTGTTGATCCACCTGTTAAACGTAA
CAGAGGAAGGCTCCTGGTTCTAGTAAGAAACAGCTTGATGCTTTAGGAGGAACCTCAGG
AGTTGGGTTTACACCTCATGTCTATTGAAGTGAACACAGGAGAGGACATAGCGTCAAAGGT
GATGGCTTTTTCGGATCAAGGGTCAAGAACAATTTGTATTCTCTCTGCAAGTGGTGCAGT
TTCTAGAGTGATGCTCTCGTCAAGCTTCTCATTCTAGTGGAATCGTTACTTATGAGGGACG
ATTTGAGATCATTACTCTCTCAGGCTCAGTCTTGAATTATGAGGTAAATGGTTCCACCAA
CAGAAGTGGTAACTTGAGTGTGGCTTTGGCTGGACCTGATGGCGGCATCGTAGGTGGCAG
TGTAGTTGGTAACTTAGTAGCTGCAACACAAGTCCAGGTGATAGTGGGAAGCTTTGTTGC
AGAAGCAAAGAAACCGAAACAAGTAGTGTTAACATTGCTCGGGGGCAGAATCCTGAACC
GGCTTCAGCGCCGGCTAACATGTTGAACCTTGGATCAGTCTCTCAAGGACCATCGAGCGA
GTCATCAGAAGAGAATGAGAGCGGTTCTCCTGCAATGCACCGTGACAATAATAATGGGAT
ATATGGAGCTCAACAACAACAACAACAACCTTCTCATCCTCATCAGATGCAAATGTA
CCAACATCTTTGGTCTAATCATGGTCAATAAATGAAGCGGAAATTAATTTGTTTCCGTT
TTGGTTACGGTTATGGTTTGATTCTT

>G1127 Amino Acid Sequence (domain in AA coordinates: 103-110, 155-162)

MDSRDIPPSHNQLQPPPGMLMSHYRNPNAASPLMVPSTSQPIQHPRLPFGNQQSQT
HQQQQQQMDQKLTLES LGFGDGSPSSQPMRFGIDDQNLQVKKRGRPRKYTPDGSIALG
LAPTSPLLSAASNSYEGGVGSDSGNGNSVDPPVKRNRGRPPGSSKKQLDALGGTSGVGF
TPHVEVNTGEDIASKVMFSDQGSRTICILSASGAVSRVMLRQASHSSGIVTYEGRFBI
ITLGSVNLNVEVNGSTNRSNLSVALAGPDGGIVGGSVVGNLVAATQVQVIVGSFVAEAK
KPKQSSVNIARGQNPASAPANMLNFGSVSQGPSSSESSEENESGSPAMHRDNNNGIYGA
QQQQQQQLPHPHQM QMYQHLWSNHGQ*

>G1131 (57..758)

TCGACTCCTCTCCTGATTGCTTCACCTTCTTCTTTACTACAGGTTTCAGCTCCTCAATGT
CCATGGATTGCTTAAGCTACTTCTTTAACTACGATCCTCCTGTCCAGCTCCAGGATTGCT
TTATTTCCGAGATGATGATTATCCCTGAAACCGATAGTTTCTTCTTCCAATCTCAAC
CGCAACTGGAGTTTCATCAGCCATTGTTCAAGAAGAAGCTCCTTCACAGACCCACTTTG
ACCCTTTCTGCGACCACTTTCTTCTCCGCAAGAAATCTTCTCCCTAACCTAAAAACG
AAATCTTCAACGAAACACAGACCTCGATTCTTCTTCTCCCAACGCAACGCGAGAGAC
TTGTTAACTCCAGCTACAATTGTAACTCAAAACCATTTCCAGAGCCGTAACCCGAATT
TCTTCGACCCCTTTCGGCGACACTGATTTCGTTCCAGAATCTTGTACCTTCCAGGAGTTTC
GAGTTCCGGATTCTCTTTAGCTTTCAAGGTAGGCCGGGGAGATCAAGATGACTCAAAGA

AGAAACAACGTAAAACCTACGCCGAAGTCCTAAGCCCGAGAGTTGTCCCGAGCCCTCGTC
CTTCACCGCCTGTTCTAAGCCCAAGAAAACCGCCTCTTAGCCCGCGCATCAACCACCACC
AGATTACCACCACCTACTTCTCCCTCCCATAAGTCCTCGAACACCTCAGCCAACAAGCC
CATACCGGGCCATTCCACCGCAACTACCACTCATCCACAGCCTCCGCTTCGCTCTTACA
GCTCATTGGCCAGTTGCAGCAGCTTAGGAGATCCACCTCCATACTCTCCTGCTTCATCTT
CTTCATCTCCTTCAGTTAGTAGTAACCATGAGAGTAGTGTGATCAATGAGCTTGTGCTA
ACTCAAAATCGGCTTTGGCTGATGTGGAAGTGAAGTTTTCAGGAGCTAACGTGCTGCTCA
AAACGGTGTGCGATAAGATCCCGGGACAAGTTATGAAGATAATTGCTGCTCTTGAAGATT
TGGCTCTTGAGATTCTTCAGGTTAATATTAACACCGTCGACGAAACCATGCTTAATTCTT
TCACCATCAAGATTGGAATTGAGTGCCAACTAAGTGCAGAAGAACTGGCTCAACAAATTC
AGCAAACATTCTGCTAGTAAAGAAGGATTAAATATAGCTTCGTATAAACCTTAACGAGAG
AGCAGTACGTACTACTTTCTCTCCTTAGTATCCCTTTAATTATCTTTTCAGTTTTCTGC
AAAGATATGGAGTTTAAAAAATAAAATTTGTTATCTAAAGTTTAAATCAAATATTGATTA
ATTATAACTAATATAGGTATAAGTGAGTTTAAAGATTATCAGCTTCATAACAGCCATCG
TCATGTTTACTTTCTTTTAAATTTTAGAATTTAGACGTACTCCTACCATGTAATTTTATT
TCTGTCTATACATCAAGCATTGTAGCTGTAATTGCATATGAATGAACAATAGTGTATGAG
TGATCTCATGAATAATATTCTTCTTGCAACACAAAAA

>G1229 Amino Acid Sequence (domain in AA coordinates: 102-160)

MQEILPDFLEECEFDVDTSLAGDDLFAILESLEGAGEISPTAASPDKDGTSSKELVKDQD
YENSSPKRKKQRLETRKEDEEEDGDGEAEEDNKQDQKQMSHVTVERNRRKQMNHILT
VLRSLMPCFYVVRGDDQASIGGVVEYISELQVQLQSLEAKKQKRTYAEVLSPRVVPSRP
SPVLSRPKPLSPRINHQIHHHLLLPPISPRTPQPTSPYRAIPPQLPLIPQPLRSYS
SLASCSLGDPPPYSPASSSSPSVSSNHSSVINELVANSKSALADVEVKFSGANVLLK
TVSHKIPQVMKIIAALEDLLEILQVNINTVDETMNSFTIKIGIECQLSAEELAQQIQ
QTFC*

>G1246 (1..1746)

ATGATCATGTACGGAGGAGGAGGAGCAGGGAAGGACGGTGGATCCACCAATCACTTATCA
GACGGAGGAGTGATATTGAAGAAAGGTCCATGGACGGCGCGGAAGATGAGATACTTGCT
GCGTACGTTAGAGAGAACGGTGAAGGGAATTGGAACGCCGTTTCAGAAAAACACAGGTTT
GCTCGTTGCGGCAAAAGCTGCCGTCTTCGATGGGCCAATCACCTCCGACCAAATCTGAAA
AAAGGCTCTTTCACCGGTGACGAAGACGTCTCATCATTCAGCTTCATGCTCAGCTTGGT
AACAAATGGGCTCGCATGGCTGCTCAGTTACCGGGAAGAACAGACAACGAGATTAAAGAAC
TATTGGAACACGAGATTGAAACGACTTCTTCGCCAAGGACTTCCTTTATCCTCCAGAT
ATTATCCCTAACCATCAACTCCATCCACATCCACATCATCAACAACAACAGCAACATAAC
CATCATCATCATCATCAACAACAACAACATCAACAATGTATTTTCAACCACAA
TCTTCACAACGAACACACCATCATCTTCCCTCTTCCATCTCCAACACCAGCAAAACGCA
AAGTCTCATCATCTTCACTTTTCATACCAGACTGCTAACCTCCTCCATCCACTTAGC
CCTCACACTCCAAACACACCATCTCAACTCTCTTCCACACCGCCTCCACCACCACTTTCC
TCTCCTTTATGTTCCCTTCGCAACAACCAATACCCGACCCTTCCCTCTTTGCCCCCCCG
CGTTCCCAAATCAACAACAACAACCGGAAATTTCACTTTCCCTAGACCTCCACCTCTC
CTTCAACCGCCTTCATCACTCTTCGCAAAACGTTACAACAATGCTAACACTCCTCTTAAT
TGCATCAACCGCTCTCAACCGCACCATTTTCCCTGTTTCAAGAGACTCCTACACTTCC
TTTCTTACATTGCGCTTACCCTTCCCAACCGCTCAAACCGCTACTTACCACAATACTAAT
AACCCTTACTCTTCTCTCTCTCTCTTCTTTAAACCTTCTTCTTCTTCTTACCCTACA
TCAACTTCTTCCCAAGCTTTCTTCACTCCATTACACTCCTTCTTCCACCTCATTTTAT
ACCAACCCAGTTTACTCCATGAAACAAGAGCAGTCCCTTCAAACCAAATCCCCAAATA
GATGGCTTCAATAACGTCAACAACCTTCACAGACAACGAGAGACAGAATCATAACCTTAAC
AGTTCCGGTGCTCAFAGAAGAAGTAGTAGCTGCAGCCTCTTAGAGGATGTCTTCGAAGAG
GCCGAAGCTTTAGCTCTGGAGGCGAGGCGGACCTCCAAAACGAAGACAACCTCACAGCT
TCTCTTCCGAACCAACAACAACAACCAACAACAACGACAACCTTCTTCTCGGTTAGTTTC
GGACATTATGATTCTTCTGACAACCTTATGTTCTTGCAGATTGAAATCAAAGGAAGAA
GAGTCTCTTCAAATGAACACAATGCAGGAGGACATAGCTAAGCTTCTTGATTGGGGAAGT
GATAGTGGAGAGATCTTAATGGACAATCATCTGTGTCACTGACGACAATCTGTTCTT
GATGTTTCATCAATTAGCTTCACTATTCCCGGTGATTCTACAGCCGTGCTAGCCGCAACA
AACGACCAACAACAAGAATAATAACAATAATTGTTCTGCGATGACATGCAGGGAATA
AGGTAG

>G1246 Amino Acid Sequence (domain in AA coordinates: 27-139)

MIMYGGGAGKDGSTNHLSDGGVILKKGPTAAEDEILAAYVRENGEGNWNNAVQKNTGL
ARCGKSCRLRWANHLRPNLKKGSFTGDEERLIIQLHAQLGNKWARMQAQLPGRTDNEIKN
YWNTRLKRLLRQGLPLYPPDIIPNHQLHPPHHQQQQQHNHHHHHHQQQQQHQQMYFQPPQ
SSQRNTPSSSPLPSPTPANAKSSSFTFHTTTANLLHPLSPHTPNTPSQLSSSTPPPPPLS
SPLCSPRNNQYPTLPLFALPRSQINNNNGNFTFPRPPPLQPPSSLFKRYNNANTPLN
CINRVSTAPFSPVSRDSYTSFLTLPYPSPTAQATYHNTNNPYSSSPSFLNPSSSSYPT
STSSPSFLSHYTPSSTSFHTNPVYSMKQEQLPSNQIPQIDGFNNVNNFTDNERQNHNLN
SSGAHRRSSSCSLEDVFEEAEALASGGRGRPPKRRQLTASLPNHNNTNNNDNFFSVSF
GHYDSSDNLCSLQDLKSKEESLQNMNTMQEDIAKLLDWGSDSGEISNGQSSVVTDDNLVL
DVHQLASLFPADSTAVVAATNDQHNKNNNNCSWDDMQGIR*

>G1255 (138..1388)

CAGCTCAAACCTCTCTAGGACTACACTAAATCTAACTTTTTGCAGAGAGCAAAAGATTCAA
TAATTGAGATTGATCTCAAAACCAAGCTCTCGTGTCTTGTCTGTTGATGTTGGTTGTGT
AGACTTTGTATACAATGATGAAAAGTTTGGCGAATGCTGTTGGAGCGAAGACGGCGAGGG
CTTGCGACAGCTGCGTGAAGAGACGTGCACGGTGGTACTGCGCGGCCGACGATGCTTTTC
TTTGCCAGTCTTGCGACAGTTTGGTCCATTACAGCAAACCTCTTGCTCGCGCCACGAGA
GAGTCCGTTTGAAGACGGCTAGCCCGCGGTCTGTAAGCATAGCAACCACTCATCAGCTT
CTCTCCACATGAGGTGCGCCAGTGGCATCACGGGTTTACTCGTAAAGCTCGAACGCCAC
GTGGCTCTGGTAAAGAAAACAATTCTGTCGATATTTTCATGACTTGGTTCCTGATATTAGTA
TTGAGGATCAGACAGACAATATGAGCTTGAAGAGCAGCTGATCTGTCAAGTGCCGGTTC
TAGATCCGTTGGTGTCTGAGCAGTTCTGAACGATGTCGTTGAGCCCAAGATCGAGTTTC
CTATGATCAGAAGTGGTTTATGATGATCGAGGAGGAGGAAGACAACGCTGAAAGTTGTCTTA
ATGGATTTTCCCGACCGACATGGAGCTTGAGGAGTTTGTCTGCTGACGTGGAGACTCTGC
TCGGTCCGGGTTAGACACGGAGTCGTATGCCATGGAGGAGCTAGGGTTATCTAATTCAG
AGATGTTCAAAATCGAAAAGATGAGATTGAAGAAGAACTAGAAGAGATAAAAGCCATGA
GCATGGATATATTGATGATGATCGAAAAGACGTGGATGGAACAGTACCGTTTGAGCTAA
GCTTTGATTACGAGTCGTACACAAGACGTCCGAAGAAGAGGTAATGAAGAAGTTGAAA
GTAGTGGTGAATGTGTTGTTAAGGTGAAGAGGAAGAACATAAGAATGTTCTGATGCTAA
GATTAAACTATGACTCGGTGATATCCACTTGGGGAGGTCAAGGTCCACCGTGGAGTTCAG
GAGAGCCACCGGAACGAGACATGGACATCAGCGGTTGGCCAGCCTTTTCCATGGTGGAGA
ATGGAGGAGAAAGTACTCATCAGAAGCAATACGTTGGTGGATGTTTACCATCAAGTGGGT
TTGGAGATGGAGGTAGAGAAGCTAGAGTTTCGAGATACAGAGAGAAGAGGAGGACAAGGT
TGTTTTCTAAGAAGATACGGTACGAGGTACGTAAATTGAATGCAGAGAAAAGACCACGAA
TGAAAGGAAGATTCTGTAAGAGAGCCTCGCTCGCTGCTGCTTACCATTAGGTGTTA
ATTACTGAATAGTTAATATCTATTATCTCACTTTACAAATTTCCGTGAATCT
TTTTTCTTCTGAAACAACAGAAGTTATTTTGGCACTTAATGTGCTTTGAGGACTTGTAT
GTACATAGAAGTAACCAATAATAATGTGACTTTTACTA

>G1255 Amino Acid Sequence (domain in aa coordinates: 18-56)
MKSLANAVGAKTARACDSCVKRRARWYCAADDAFLCQSCDSLVSANPLARRHERVRLKT
ASPAVVKHSNHSSASPPHEVATWHHGFTRKARTPRGSGKKNSSIFHDLVPDISIEDQTD
NYELEEQLICQVPVLDPLVSEQFLNDVVEPKIEFPMIRSGLMIEEEDNAESCLNGFFPT
DMELEEFADVETLLGRGLDTSYAMEELGLSNSEMFKIEKDEIEEEVEBIKAMSMDIFD
DDRKDVDGTVPFELSFDYESHKTSEEEVMKNVSSGECVVKVKEEHNKVMMLRLNYDS
VISTWGGQGGPPWSSGEPPEPDMDISGWPAPFSMVENGGESTHQKQYVGGCLPSSGFGDGG
EARVSRVREKRRLTFSKKIRYEVRLNAEKRPRMKGRFVKRASLAAAASPLGVNY*

>G1304 (1..978)

ATGGGGCGATCACCATTGTCGATGAGAATGGTCTAAAGAAAGGGCCATGGACACAAGAG
GAGGATGATAAACTGATAGATCACATTCAAAAACATGGCCATGGCAGCTGGAGAGCTCTT
CCAAAGCAAGCCGGTTTAAACCGATGCGGAAAGAGTTGTAGATTAAAGATGGACCAACTAC
TTGAGACCTGACATCAAGAGAGGAAATTTCACTGAAGAGGAAGAACAACTATTATCAAC
CTCCATTCCCTTCTTGGAACAAGTGGTTCGTCGATAGCCGGTAATCTTCTGGAAGAACG
GACAATGAAATAAAAACTATTGGAACACACATTTGAGAAAGAACTTCTCAAATGGGG
ATTGATCCGGTGACCCATAGGCAAGAACCGACCATTAACGTTTATAGCAGCTCTCCCG
CAGCTTATAGCCCGCGCAAAATTTCAACAGCCTCTTGAATCTCAACCAAAATGTGCAACTG
GATGCAACAACCTTGTCTAAAGCTCAACTGCTACACACTATGATTCAAGTCCTTAGCACC
AATAACAACACCACCAATCCTTCTTTTCTTATCAACTATGCAAAACAGTAACACCAAT
CTCTTTGGCCAAGCTTCTTACTTAGAGAACCAAAATCTTTTGGTCAGTCTCAAACTTC

TCTCACATTCTTGAGGATGAGAATTTGATGGTCAAACCCAAATTATTGATAACCCCTTG
GACTCTTTTCTTCCCCCATACAACCCGGTTTTCAAGATGATCATAATTCACCTCCCTCTA
TTGGTTCGCGCTCTCCTGAAGAATCTAAAGAACTCAAAGGATGATCAAGAACAAGAC
ATCGTCGATACCATCATCATGATGCTTCAAACCCCTCATCATCAAACTCAACGTTTACA
CAAGATCATCATCACCCATGGTGTGACACTATTGATGATGGAGCAAGTGATTCTTTTGG
AAAGAGATAATAGAGTAA

>G1304 Amino Acid Sequence (conserved domain in AA coordinates:13-118)

MGRSPCCDENGLKKGPTWQEEEDDKLIDHIQKHGHSWRALPKQAGLNRCGKSCRLRWNY
LRPDIKRGNFTEEEQTIINLHSLGKNWSSIAGNLPGRDNEIKNYWNTHLRKKLLQMG
IDPVTHRPRDHLNVLAAALPQLIAANFNSSLNQLNQNVLQDATTAKAQLLHTMIQVLST
NNNTTNPSFSSSTMQNSNTNLFQASYLENQNLFGQSQNFHILEDENLMVKTQIIDNPL
DSFSSPIQPGFQDDHNSLPPLVPASPEESKETQRMINKNDIVDYHHHDASNPSSNSTFT
QDHHHPWCDTIDDGASDSFWKEIIE*

>G1318 (7..849)

AAAAATATGAGGAAGCCAGAGGTAGCCATTGCAGCTAGTACTACCAAGTAAAGAAGATG
AAGAAGGGACTTTGGTCTCCTGAGGAAGACTCAAAGCTGATGCAATACATGTTAAGCAAT
GGACAAGGATGTTGGAGTGATGTTGCGAAAAACGCAGGACTTCAAAGATGTGGCAAAAGC
TGCCGTCTTCGTTGGATCAACTATCTTCGTCTGACCTCAAGCGTGGCGCTTCTCTCCT
CAAGAAGAGGATCTCATCATTCGCTTTCATTCATCCTCGGCAACAGGTGGTCTCAGATT
GCAGCACGATTGCCTGGTTCGGACCGATAACGAGATCAAGAATTTCTGGAACCAACAATA
AAGAAAAGGCTAAAGAAGATGTCCGATACCTCCAACCTAATCAACAACCTCATCCTCATCA
CCCAACACAGCAAGCGATTCTCTTCTAATCCGCATCTTCTTTGGATATTAAAGACATT
ATAGGAAGCTTCATGTCTTACAAGAACAAGGCTTCGTCAACCCCTTCCTTGACCCACATA
CAAAACCAATCCATTTCCAACGGGAAACATGATCAGCCACCCGTGCAATGACGATTTT
ACCCCTTATGTAGATGGTATCTATGGAGTAAACGCAGGGGTACAAGGGGAACCTTACTTC
CCACCTTTGGAATGTGAAGAAGGTGATTGGTACAATGCAATATAAACAACCACTTAGAC
GAGTTGAACACTAATGGATCCGGAACGCACCTGAGGGTATGAGACCAGTGAAGAATTT
TGGGACCTTGACCAGTTGATGAACACTGAGGTTCTTCGTTTTACTTCAACTTCAAACAA
AGCATATGAATATTTTACGTATCTTATCTTTTTTCTATTGCGGTTTATACTCAAGAT
TCTTAGCCACACACATAAATGCAAATATATATACATTGTTAGAGAGTATTTTGTATTT
CGTATAATCTTTTCGTACTAGGGCTTGAGCCTTGAGGTCCCATGTAACGATTAGTCAATG
TAAACATATATCTATAATAAATAAAATAAAAGAAATAATAAGCACATAAAAAAAAAAA
A

>G1318 Amino Acid Sequence (domain in AA coordinates: 20-123)

MRKPEVAIAASTHQVKMKKGLWSPEEDSKLMQYMLNQGQCSWDVAKNAGLQRCGKSCR
LRWINYLRLPDLKRGAFSPQEEEDLIIRFHSILGNRWSQIAARLPGRDNEIKNFWNSTIKK
RLKKMSDTSNLINNSSSSPNTASDSSSNSASSLDIKDIIGSFMSLQEQGFVNPSLTHIQ
NNPFTGNMISHPCNDDFTPYVDGIYGVNAGVQGEYFPFPLECEEGDWNANINNHDEL
NTNGSGNAPEGMRPVEEFWDLQMLNTEVPSFYFNFKQSI*

>G1320 (39..788)

GAAGATCATAAAGATCAAAGGAGAGAGGTATTAATAAATGATGTGTAGTCGAGGCCATT
GGAGACCTGCAGAAGACGAGAAGCTAAGAGAACTCGTCGAGCAATTTGGTCTCATAATT
GGAACGCCATAGCTCAGAAGCTCTCTGGTTCGATCTGGTAAGAGTTGTAGATTGAGATGGT
TTAATCAATTGGATCCTAGGATTAACCGAAACCTTTACGGAGGAAGAAGAAGAAAGGC
TTTTAGCGCTCATCGGATCCATGGGAACAGATGGTCTGTGATCGCTAGATTTTTTCCCG
GTCGAACGTATAACGCTGTATAAAACCATTGGCACGTCATCATGGCTCGTCGTGGCCGAG
AACGGTCCAAGCTCCGTCCACGAGGCCCTTGGCCATGATGGCACGGTGGCTGCGACTGGGA
TGATTGGTAATTATAAAGACTGCGATAAGGAGAGAAGATTGGCAACCACAACCGCTATCA
ATTTTCCTTATCAATTCTCTCATATTAATCATTTTTCAAGTCCTCAAAGAGTCCTTGACCG
GAAAGATCGGGTTCAGAAATAGTACTACTCCAATACAAGAAGGAGCAATAGACCAACTA
AACGACCGATGGAGTTCTACAATTTTCTCCAAGTAAACACGGATTGGAAGATACACGAAT
TGATAGATAATTCAAGAAAAGACGAAGAAGAAGATGTCGATCAAAACAACCGAATTCGTA
ACGAGAATTGTGTCCATTTTTCGACTTTTTGTCTGTGGAAACTCTGCCTCTCAGGGTT
TATGTTAATTTGTCCGTACCATGTACTATAAGGTGGACCATATGTTAACTAAAGATAA
TGTAAGAAAGTACTAATCAATTAGAGCTCTGTTTGAGCCAAATGTGAAAATTAGTTAAGA
CATCCCAACATTTTTCTGTATAACACATATAAGGTTGTACTTTTATCAGGTCTAATTTT
CTATTTTATTTTAAAGGATGTTTAAATCAGACCCATAACCATTCGATAAAAAAAAAAAAAA

>G1320 Amino Acid Sequence (domain in AA coordinates: 5-108)
MMCSRGRHWRPAEDEKLRELVEQFGPHNWNIAQKLSGRSGKSCRLRWFNQLDPRINRNP
TEEEEEERLLAPHRIHGNRWSVIARFFPGRTDNAVKNHWHVIMARRGRERSKLRPRGLGHD
GTVAATGMIGNYKCDKERRLATTTAINFPYQFSHINHFOVLKESLTGKIGFRNSTTPIQ
EGAIDQTKRPMFYNFLQVNTDSKIHOLIDNSRKDEEEDVDQNNRIRNENCVPFFDFLSV
GNSASQGLC*

>G1330 (36..959)

GTACCGGCGACCTCTTTGTGGGTCACTCTTCATCAATGGGTGACAAAGGAAGGAGCTTAA
AGATCAACAAGAACATGGAGGAATTCACGAAAGTGGAAGAAGAAATGGACGTAAGGAGAG
GTCCATGGACAGTTGAGGAAGATTTAGAGCTCATCAATTACATTGCTAGTCATGGTGAAG
GTCGATGGAACTCTCTCGCTCGTTGCCGGAACCTCAAAGGACCGGAAAAAGCTGCAGAC
TTCGGTGGCTGAACATCTCCGACCAGATGTGCGCCGTGGAAACATAACCTCGAAGAAC
AACTCTTGATTCTTGAACCTCACACACGTTGGGGCAATAGATGGTCTAAGATTGCACAAT
ATTTACCAGGAAGAACGATAACGAGATCAAAAACCTATTGGAGAACACGTGTTCAAAAGC
ATGCAAAACAGCTTAAATGCGACGTTGACAGTCAACAATTTAAAGACACCATGAAGTATC
TTTGGATGCGCTCGGCTCGTAGAAAGGATCCAAGCCGCGTCCATCGGGTCTGTTTCCATGT
CATCTTGCCTACACCTCCTCAGATCAGTTCGTGATCAACAACAACACCAACAACG
TGGATAATTTGGCTTTTAAATGAGTAACCTAATGGTTACATCAGCCGGATAATTCCAGCG
TGGCAGTATCTCCTGTATCAGATTTGACGGAGTGTCAAGTGAGTAGTGAAGTGTGGAAGA
TTGGTCAGGATGAGAATTTGGTGGATCCAAAATGACATCGCCGAATTATATGGATAATA
GCAGTGGACTATTAAACGGAGATTTTACGAAGATGCAAGATCAAAGTGACCTTAATTGGT
TTGAAAATATTAATGGGATGGTACCAATTTATTCGGACAGTTTTTGAACATTGGAATG
ATGAAGACTTCTGGCTCTTACAACAACATCAACAAGTCCACGACAATGGAAGCTTCTGAA
TAGACAAGAAGCTATGCGGCC

>G1330 Amino Acid Sequence (domain in AA coordinates: 28-134)
MGDKGRSLKINKNMEEFKVEEEMDVRRGPWTVEEDLELINYIASHGEGRWNSLARCAEL
KRTGKSCRLRWLNLYRPDVRGNITLLEQLLILELHTRWGNRWSKIAQYLPGRDNEIKN
YWRTRVQKHAKQLKCDVNSQQFKDTMKYLWMPRLVERIQAASIGSVSMSSCVTTSSDQFV
INNNNTNNVDNLALMSNPNGYITPDNSSVAVSPVSDLTECQVSSEVWKIGQDENLVDPKM
TSPNYMDNSSGLNNGDFTKMQDQSDLNWFENINGMPVNYSDSFWNIGNDEDFWLLQQHQ
VHDNGSF*

>G1352 (79..900)

GCGCGATTAAAACTCTCAACTTTTCTCTCAAATTTCTGATCCTTTGATCCAACAGTTAG
AAGAAGATTTCATCTGATCATGGCCCTCGAAGCGATGAACACTCCAACCTCTTCTTTTACC
AGAATCGAAACGAAAGAAGATTTGATGAACGACGCCGTTTTCATTGAGCCGTGGCTTAAA
CGCAAACGCTCCAAACGTCAGCGTTCTCACAGCCCTTCTCGTCTTCTTCTCACCAGCT
CGATCTCGACCCAAATCCAGAATCAAGATCTTACGGAAGAAGAGTATCTCGCTCTTTGT
CTCCTCATGCTCGCTAAAGATCAACCGTCGCAACGCGATTTCATCAACAGTCGCAATCG
TTAACGCCGCCGCCGAGATCAAGAACCTTCCGTACAAGTGTAACGCTGTGTAAGAAAGCG
TTTCTCTCTATCAGGCTTTAGGCGGTACAAAGCAAGTCACCGAATCAAACCACCAACC
GTAATCTCAACAACCGCCGATGATTCAACAGCTCCGACCATCTCCATCGTCGCCGGAGAA
AAACATCCGATTGCTGCCTCCGGAAGATCCACGAGTGTCAATCTGTCAATAAGTGTTT
CCGACGGGTCAAGCTTTAGGCGGTACAAACGTTGTCACTACGAAGGCAACCTCGGCGGC
GGAGGAGGAGGAGGAAGCAAATCAATCAGTCACAGTGGAGCGTGTCCGAGCACGGTATCG
GAAGAAAGGAGCCACCGTGGATTTCATCGATCTAAACCTACCGCGTTACCTGAACTCAGC
CTTCATCAACAATCCATCGTCGACGAAGAGATCTTGAGTCCGTTGACCGGTAAAAAACCG
CTTTGTGTGACCGATCACGACCAAGTCATCAAGAAAGAAGATTTATCTTTAAAAATCTAA
TACTCGACTATTAAFTCTTGTGTGATTTTTTTCGTTACAACCATAGTTTCATTTTCATTT
TTTTAGTTACAAATTTTAAATTGTTCTGATTGATTGAAA

>G1352 Amino Acid Sequence (domain in AA coordinates: 108-129,167-188)

MALEAMNTPSSFTRIETKEDLMNDAVFIEPWLKRKRKRQRSHSPSSSSSSPPSRPKS
QNQDLTEEEYLALCLLMLAKDQPSQTRFHQQSQSLTPPPESKNLPYKCNVCEKAFPSYQA
LGGHKASHRIKPTPTVISTTADDSTAPTISIVAGEKHPAASGKIHECSICHKVFPTGQAL
GGHKRCHYEGNLGGGGGGGSKSISHSGSVSSTVSEERSHRGFIDLNLPALPELSLHNPI
VDEEILSPLTGKKPLLLTDHDQVIKKEDLSLKI*

>G1354 (1..1047)

ATGGAAAGTCTCGCACACATTCCTCCCGTTATCGATTCCATCCGACCGATGAAGAACTC

GTTGACTATTATCTCAAGAACAAAGTTGCATTCCCGGGAATGCAAGTTGATGTTATCAAA
GATGTTGATCTCTACAAAATCGAGCCATGGGACATCCAAGAGTTATGTGGAAGAGGGACA
GGAGAAGAGAGGGAATGGTATTTCTTTAGCCACAAGGACAAGAAATATCCAAGTGGGACA
CGAACCAATAGAGCAACGGGCTCCGGATTTTGGAAGCAACGGGTCGAGACAAGGCCATT
TACTCAAAGCAAGAGCTTGTGTGGGATGAGGAAGACTCTTGTCTTTTACAAAGGTAGGGCC
CCAAATGGTCAGAAATCTGATTGGATAATGCACGAATACCGTCTTGAGACCGATGAAAAT
GGACCGCCTCATGAGGAAGGATGGGTGGTTGTGCGCGCTTCAAGAAGAAGCTAACCACG
ATGAACTACAACATCCAAGAACAATGATGGGATCATCATCAGGCCAAGAATCTAAGTGG
TTCACGCAGCAAATGGATGTGGGGAATGGTAATTACTATCATCTTCTGTATCTAGAGAGT
CCGAGAATGTTTCAAGGCTCATCATCATCACTATCATCATTACATCAGAATGATCAA
GACCCCTTATGGTGTCTGACTCAGCACTATTAACGCAACCCCAACTACAATAATGCAACGA
GATGATGGTCATGTGATTACCAATGATGATGATCATATGATCATGATGAACACAAGTACT
GGTATCATCATCAATCAGGATTACTAGTCAATGATGATCATAATGATCAAGTAATGGAT
TGGCAAACGCTTGACAAGTTTGTGCTTCTCAGCTAATCATGAGCCAAGAAGAGGAAGAA
GTTAACAAGATCCATCAGATAATTCTTCGAATGAAACATTTTCATCATCTCTCTGAAGAG
CAAGCTGCAACAATGGTTTCGATGAATGCTTCTTCTCTCTCTCCATGTTCTCTTCTAC
TCTTGGGCTCAAAAATACACACAGTAA

>G1354 Amino Acid Sequence (domain in AA coordinates: TBD)
MESLAHIPPGYRFHPTDEELVDYLLKNKVAFFPMQVDVIKVDLYKIEPWDIQELCGRGT
GEEREWYFFSHKDKKYPTGTRNRTGSGFWKATGRDKAIYSKQELVGMKRLTVFYKGRA
PNGQKSDWIMHEVRLETDENGPPHEEGWVVCRAFKKKLTMTNYYNNPRTMMGSSSGQESNW
FTQQMDVGNNGNYHLPDLESPRMFQSSSSSLSLHQNDQDPYGVVLSINATPTTIMQR
DDGHVITNDDHMMIMNTSTGDHHQSGLLVNDHDNDQVMDWQTLDFVASQLIMSQEEEE
VNKDPDSNSSNETFHLSEEQAATMVSMNASSSSSPCSFYSWAQNTHT*

>G1360 (1..1257)

ATGGGAGATAGAAACAACGACGGTGATCAGAAAATGGAGGATGTATTGTTGCCCGGATTT
AGGTTTCATCCAACCGACGAAGAGCTCGTAAGCTTCTACCTGAAGCGGAAGGTTCAACAC
AACCCTCTCTCCATGAGCTCATAAGACAACCTCGATATCTACAAATATGACCCCTGGGAT
CTTCCAAAGTTTGCGATGACGGGTGAAAAAGAATGGTACTTTTATTGTCCAAAGGACAGG
AAGTATAGGAACAGCTCGAGGCCAAACCGAGTGACCGGAGCTGGTTTTTGGAAGCCACG
GGAACGGACCGCCGATATACTCGTCAGAAGGAAACAATGCATAGGTTTAAAGAAGTCC
TTAGTGTTCTACAAAGGAAGAGCAGCGAAAGGAGTTAAGACTGATTGGATGATGATGAG
TTTCGTTTGCTTCTCTCTCCGAACCATCTCTCTCTTCTAAGAGATTCTTCGACTCTCCT
GTCTCTCCCAACGATTCTATGGGCTATATGCAGAATCTTCAAAAAGACCAACACAACGACC
CTAAGAGCTCTCTCTCACTCTTTTGTTCCTCGTTACCACCAGAAACAAGCACCGACACA
ATGTCTAACCAAAAGCAATCAACACATACCATTTTTCTTCAGACAAGATCCTCAAACCT
AGCTCTCACTTCCAGTTTACCATGAGAATATGAACACTCCCAAACTAGTAATAGTACA
ACTCCATCCGTTCCCACTATAAGTCCCTTCTTACTTGGATTTCACCTTCATACGACAAA
CCCACCAACGTTTTCAATCCGGTTTCATGTTTACACCAACAATACCTCACAAATCTCTTT
CTTGCCACACAAGAAACACAACCTCAGTTTCCAGGCTCCCTCGTCAAATGAAATCCCA
TCGTTTCTGCTAAACACGTCTTCAGATTGACCTTCTTGGGAGAATTCACGAGCCATATC
GACCTCAGCGCAGTGTGGCCCCAAGAGCAATGTCCCCCGCTTGTAAGCCTACCACAGGAG
TATCAAGAGACGGGATTGCAAGGAAATGGTATAATGAAGAACATGCGTGGTTCCAATGAA
GATCATCTTGGTGATCATTGCGACACACTTCGGTTTGATGATTTCACTTCAACAATTAAT
GAGAACCATCGTCATCATCAAGACCTGAAACAGAACATGACATTGCTGGAGAGTTATTAT
TCTTCTTTATCGTCCATCAATAGCGATTGTCAGCTTGTCTTCTCCAGTACAACCTGA

>G1360 Amino Acid Sequence (conserved domain in AA coordinates:18-174)

MGDRNNDGDQKMEDVLLPGFRFHPTDEELVSFYLLKRVQHNPLSIELIRQLDIKYDPWD
LPKFAMTGEKEWYFYCPRDRKYRNSRPNRVTGAGFWKATGTDRIYSSEGNKICIGLKK
LVFYKGRAAKGVKTDWMMHEFRPLSLSESPSPSKRFFDSPVSPNDSWAICRIFKKTNTTT
LRALSHSFVSSLPETSTDMSNQKQSNTHYFSSDKILKPSSHQFHHENMNTPKTSNST
TPSVPTISPFSTYDPTSYDKPTNVFNPVSLDQQYLTLNLFATQETQPPQFRLPSSNEIP
SFLNLTSSDSTFLGEFTSHIDLAVLAQEQCPPLVSLPQEQETGFEGNGIMKNMRGSNE
DHLGDHCDTLRFDDFTSTINENHRHHQDLKQNMTLLESYSSSLSSINSDLPACFSSTT*

>G1364 (1..537)

ATGGCGGAGTCGCAAGGCAAGAGTCCCGGAGGCTGTGGAAGCCATGAGAGTGGTGGAGAT
CAAAGTCCAGGTCGTTACATGTTCTGTCGAGCAAGATAGGTTTCTTCCGATTGCTAACATA

AGCCGTATCATGAAAAGAGGTCTTCTGCTAATGGGAAAATCGCTAAAGATGCTAAGGAG
ATTGTGCAGGAATGTGTCTCTGAATTCATCAGTTTCGTCAACGAGCGAAGCGAGTGATAAA
TGTCAAAGAGAGAAAAGGAAGACTATTAATGGAGATGATTGCTTTGGGCAATGGCTACT
TTAGGATTTGAAGACTACATGGAACTCTCAAGGTTTACCTGATGAGATATAGAGAGGGT
GACACAAAGGGATCAGCAAAGGTGGGGATCCAAATGCAAAGAAAGATGGGCAATCAAGC
CAAAATGGCCAGTTCTCGCAGCTTGCTCACCAAGGTCTTATGGGAACCTCTCAAGTAAC
TTTCTCTCTCTCTCTTCACTCAAGCAATACGCATCATTCTCTTCTAATTGTAA
>G1364 Amino Acid Sequence (conserved domain in AA coordinates: 29-120)
MAESQAKSPGGCGSHESGGDQSPRSLHVREQDRFLPIANISRIMKRLPANGKIAKDAKE
IVQECVSEFISFVTSEASDKCQREKRKTINGDDLWAMATLGFEDYMEPLKVYLMRYREG
DTKGSAGGDPNAKKGQSSQNGQFSQLAHQGPYGNISQVTFPLFSSSHSSNTHHSLIC*
>G1379 (68..622)
CTCTGCCTCTCTCTCTCTCAAAACCATCTCGAAAGTCTTCTCTTTTCGAGGGTTTAG
ATCCTCCATGGAAGCGCGGAGTTGCTGACGTGGCTGTCCCGGTACGAGGAAGAGAGA
CAGACCTTACAAAGGAATTAGGATGAGGAAGTGGGAAAGTGGGTGGCGGAGATTCTGTGA
GCCTAACAGCGCTCTAGGTTATGGCTTGGCTCTTACTCTACTCCCGAGGCGGCGCGCG
AGCTTACGACACGGCGGTTTTCTATCTTAGAGGACCTACGGCGAGGCTTAACCTCCCTGA
GCTTCTTCTGGGGAGAAATTTCTCCGACGAGGATATGTCGGCTGCGACCATCAGGAAGAA
AGCCACGGAGGTGGTGTCTCAGGTTGATGCTTTGGGCACGGCGGTGCAAAATAACCGCCA
CCGTGTTTTTGGTCAGAATCGAGATAGTGTGGATAATAAGAAATTTTCATCGGAATTA
TCAAACGGTGAACGAGAAGAAGAAGAAGATGAGGATGACAAGAGATTGAGGAGTGG
CGGCCGGTTATTGGATCGGGTTGACTTGAATAAATTACCCGACCCGGAAGCTCCGATGA
AGAATGGGAAAGCAACATTAATAATATATAGTTTGGAGCGGTGGCTGTTGCTAACGTAC
GCCAACGGCTTGTCTTACGAATCATTAGCGCGTTTATGATTTTTTTTTTTTTTTTTTTT
CATTATCTGAAAATTTAGGGCTTTTTAGTTATTAATTTTTGTTTTGTTTTTCTTTCT
TGCGAGTTTTCGGTTTATGGAATTTTAGGCTATTGCTTAACGAAAAAAAAAAAAAAAAA
>G1379 Amino Acid Sequence (domain in AA coordinates: 18-85)
MEGGGVADVAVPGTRKRDRPYKGIIRMKWGWVAEIREPNKRSRLWLGSYSTPEAAARAY
DTAVFYLRGPTARLNFPELLPGEKFSDEDMSAATIRKKATEVGAQVDALGTAVQNNRHRV
FGQNRSDVDNKNFHRNYQNGEREEEEDEDDKRLRSRGRLLDRLNKLDPPESSDEEW
ESKH*
>G1384 (33..977)
GTACATTTTTTTTTTGTTATTTTCAAGAACTCCGATGGCGGATCTCTTCGGTGGTGGCCACG
GCGGCGAGCTTATGGAAGCACTTCAACCTTTTTACAAAAGTGCTTCCACGTCTGCTCAA
ATCCTGCGTTTGCCTCCTCAAACGATGCGTTTGCCTTGCCTCCAAACGACCTATTTCTT
CTTCTTCTTACTATAATCCTCATGCATCTTATTCCCTTACATTCACAAACCTCTTACC
CGGATATTTATTCTGGATCCATGACCTATCCATCTTCACTTCGGGTTCGGATCTTCAACAAC
CCGAAACTACCAATCTCAGTTCCATTACCAAAACACTATCACTTACACTCACCAGACA
ACAAACACTTGCATGCTTAACTTCAATTGAGCCGAGCCAACCGGGTTTTATGACCCAACCGG
GTCCGAGTTCGGGTTTCGGTTTCAAACCGGCTAAGCTCTATAGAGGAGTGAGGCAAAGAC
ATTGGGGAAAATGGGTCGCGGAGATCCGTTTACCCAGGAACCGAACCGGACTTTGGCTCG
GAACATTGCACACGGCTGAAGAAGCCGCTTGGCTTATGATCGCGCCGCTTTAAGCTTC
GTGGTGACTCGGCTCGGCTTAACTTCCCAGCTCTCCGATACCAAACCGGCTCGTCTCCGT
CTGATACCGGCGAATATGGTCTTATTCAAGCTGCCGTAGACGCTAAACTAGAAGCCATAT
TAGCTGAGCCGAAGAATCAGCCGGGCAAAACGGAGAGGACGTCGAGGAAACGAGCTAAAG
CCCGGCTTCTTCAGCTGAGCAGCCGCTCAGCGCCACAACAACATTCCGGGTTCGGGTGAAA
GTGATGGGTTCGGGTTACCGACTTTCGGATGTTATGGTGCAGGAGATGTGCCAAGAGCCAG
AGATGCCATGGAATGAAAATTTATGCTCGGCAAGTGTCTTCTTATGAGATAGATTGGG
CTTCAATTTTATCGTGAAAAATTAGGATTCAATTCATTTTATTCAATTTAACTTGTGTTG
TATTTTCTTTTAACTTTAGGGTTATTAGCTGTGCGTAA
>G1384 Amino Acid Sequence (domain in AA coordinates: TBD)
MADLFGGGHGGELMEALQPFYKSASTSASNPFASSNDFAFASPNDLFSSSSSYNPHASL
FPSSHSTTSYPIYSGSMTYPSFSGSDLQQPENYQSQFHYQNTITYTHQDNNTCMLNFIEP
SQPGFMTQPGPSSGSVSKPAKLYRGVQRHWGKWVAEIRLPNRNTRLWLGTFTDAAEEAAL
AYDRAAFKLRGDSARLNFALRYQTGSSPSDTGEYGPIQAAVDKLEAILAEPKNQPGKT
ERTSRKRAKAAASSAEQPSAPQQHSGSGESDGSPTSVMVQEMCQEPMPWNENFMLG
KCPSEIDWASILS*

>G1399 (261..1475)

AGGTCGAATTTTCTGAAATTAAGATTCATTCTCCATGGAAGAAGCTCTGTTTTATTCT
CTTTAGCTTAGCTTAGCTTCTACTGATCTGTTTTTGCTACAAAATCCCATCTTTTCTTT
AAAACCTCTTATCTCTGAATCTTGAGTTTCTTGTAAGAAGAAGCAATTTTGAATCTTT
CGTAATCATAAAGATTCTGAGGATCTCTACTGATTGTGCGGAATCTCTCACTACAGAA
TCACCTTGATCTTATGTCCGGATGGAGGAGAGAGAAGGAACCAACATCAACAACACATCA
CTAGCAGTTTCGGCTTGAAGCAGCAACATGAAGCTGCTGCTTCTGATGGTGGTTACTCAA
TGGACCCACCACCAAGACCCGAAACCCCTAACCCGTTTTTAGTCCCACCCACTACTGTCC
CCGCGGCCGCCACCGTAGCAGCAGCTGTTACTGAGAATGCGGCTACTCCGTTTAGCTTAA
CAATGCCGACGGAGAACACTTCAGCTGAGCAGCTGAAAAAGAAGAGAGGTAGGCCGAGAA
AGTATAATCCCGATGGGACTCTTGTCTGACTTTATCGCCGATGCCAATCTCGTCTCTG
TTCCGTTGACGTCCGAGTTTCTCCAAGGAAACGAGGAAGAGGACGTGGCAAGTCTAATC
GATGGCTCAAGAAGTCTCAAATGTTCCAATTCGATAGAAGTCTCTGTGATACCAATTTGG
CAGGTGTAGGAACGTGATTTTGTGGTGCCAACTTTACACCTCATGTACTGATCGTCA
ACGCCGAGAGGATGTGACGATGAAGATAATGACATTCTCTCAACAAGGATCTCGTGCTA
TCTGCATCCTTTAGCTAATGGTCCCATCTCCAATGTTACGCTTCGTCAATCTATGACAT
CCGGTGGTACTCTAATCTATGAGGGTCGTTTGTAGATTCTCTCTTTGACGGGTTCTGTTA
TGCAAAATGACTCTGGAGGAACCTGAAGTAGAGCTGGTGGTATGAGTGTGCTTGCCTG
GACCAGATGGTCTGTCTTTGGTGGAGGACTCGTGGTCTCTTTCTTGCTGCTGGTCTG
TCCAGGTAATGGTAGGACTTTTATAGCTGGTCAAGAGCAGTCAAGCTGGAGCTAGCAA
AAGAAAGACGGCTAAGATTTGGGGCTCAACCATCTTCTATCTCCTTTAACATATCCGCAG
AAGAACGGAAGGCGAGATTCGAGAGGCTTAACAAGTCTGTTGCTATTCTGACCAACCA
CTTCATACACGCATGTAAACACACAAATGCGGTTACAGTTACTATACAACTCGGTTA
ACCATGTCAAGGATCCCTTCTCGTCTATCCAGTAGGAGGAGGAGGAGGTGAGAGGTAG
GAGAAGAAGAGGGTGAAGAAGATGATGATGAATTAGAAGGTGAAGACGAAGAATTCCGAG
GCGATAGCCAATCTGACAACGAGATTCCGAGCTGATGATGATCATAACGTTTCTTTTCGC
GGATTGTTAGGTTGATGGATTTCAGATTTTGGTGTGATTGTTTTATTAAACAGAAATG
TTTGAAGCTGCTATCTTTAGGTTCCCATCTCTTGTGATTGTTGAGTATCCTTGTTAGA
AACAACTTACTGTTGCAAACTCTCTTCAAAAAGTTTCACTTTGCTTTCCCA

>G1399 Amino Acid Sequence (domain in AA coordinates: 86-93)
MEEREGTNINNNITSSFLKQHEAAASDGGYSMDPPPRPNPNPFLVPPTTVPAATVA
AAVTENAATPFSLTMTENTSABQLKKRGRPRKYNPDGLVVTLSMPPISSSVPLTSEF
PPRKRGRGRGKSNRWLKSQMFQFDRSPVDNLAVGTADFVGANFTPHVLIVNAGEDVT
MKIMTFSQGSRAICILSANGPISNVTLRQSMSTSGGTLTYEGRFEILSLTGSFMQNDSSG
TRSRAGMSVCLAGPDGRVFGGLAGLFLAAGPVQVMVGTFIAGQEQSLELAKERRLR
GAQPSSISFNISAEERKARFERLNKSVAIAPPTTSYTHVNTTNAVHSYYTNSVNVHKDPF
SSIPVGGGGGGEVEEGEEDDELEGEDEFEFGGDSQSDNEIPS*

>G1415 (60..680)

CCTTATCACTACCAAAAGTCGTACATAATATCACTTTTCGAGTTATCAACATCCGTACA
TGTCATCCATAGAGCCAAAAGTAATGATGGTTGGTGCTAATAAGAAACAACGAACCGTCC
AAGCTAGTTTCGAGGAAAGTTGTATGAGAGGAAAAGGTGGACCCGATAACGCGTCTTGCA
CTTACAAAGGTGTTAGACAACGCACCTTGGGGCAAATGGGTGCGTGGATCCGCGAGCCTA
ACCGAGGAGCTCGTCTTTGGCTCGGTACCTTCGACACCTCCCGTGAAGCTGCCTTGGCTT
ATGACTCCGCACTCGTAAGCTCTATGGGCCTGAGGCTCATCTCAACCTCCCTGAGTCCT
TAAGAAGTTACCTAAAACGGCGTCTGCTCCGGCGTCCCAGACTACACCAAGCAGCAACA
CCGGTGGAAAAGCAGCAGCGACTCTGAGTCGCCGTGTTTCATCAACGAGATGTCATCAT
GTGGAAGAGTGACAGAGGAGATATCATGGGAGCATATAACGTGGATTGCGCGGTAATGG
ATGATTCTTCAATATGGGAAGAAGCTACAATGTCTGTTAGGATTTCCATGGGTTTCATGAAG
GAGATAATGATATTTCTCGTTTGATACTTGTATTTCCGGTGGCTATTCTAATTGGGATT
CCTTTTCACTCCCACTTTGAGGTGCTACTAGACTCTCTTAATTGTTAAGTTATCATATA
CAAACATACATATATACAAATATAGTCACCGTGAACCTAGGATATATATGTAATAAACA
CCAGTTACATGTAATATATATGTGCACATCTATATATGTGGTTTGTCTGTATAGTGTGA
AAGCAGATTCTTACCATATCA

>G1415 Amino Acid Sequence (domain in AA coordinates: TBD)
MSSIEPKVMVMVANKKQRTVQASSRKGCMRGKGGPDNASCTYKGVQRQTWGWVAEIREP
NRGARLWLTGTFDTSREAAALAYDSAARKLYGPEAHLNLPESLRSPKTAASSPASQTPSSN
TGGKSSSDSESPCSSNEMSSCGRVTEEISWEHINVLDLPVMDSSIWEEATMSLGFVWVHE

GDNDISRFDTCISGGYSNWDSFHSPL*

>G1417 (32..1501)

TCATCTCTATCTATCTCTCTTTGTCTGCAAATGGAAGAACATATTCAAGATCGCCGTGA
AATTGCGTTCTTACTACTCAGGAGAATTTCTCCACGGAGATTCTGACTCAAAGGATCATCA
ACCGAACGAGTCTCCGGTGGAAACGTCATCACGAGTCGTCTATCAAAGAAGTTGATTTCTT
CGCTGCTAAAAGTCAGCCGTTTGTATCTTGGTCATGTGAGAACAACGACGATCGTTGGATC
ATCTGGTTTTAATGATGGATTAGGTTTGGTAAATTCATGTCTATGGAACATCAAGCAATGA
TGGCGATGACAAAACCAAACCTCAAATTAGTAGACTGAAGTTGGAGCTAGAGAGGCTTCA
CGAGGAGAATCACAACTGAAGCATTATTAGATGAGGTCAGTGAGAGTTACAACGACCT
CCAAAGAAGAGTTTTGTAGCAAGACAAACACAAGTGGAAAGGTTTCATCATAAACAACA
TGAGGATGTACCTCAAGCTGGTTCTCTACAAGCTCTAGAGAACAGAAGACCAAAGGATAT
GAACCATGAAACTCCGGCCACCACCTTGAAACGACGGTCTCCAGACGACGTGGATGGTCG
TGATATGCACCGAGGATCACCAAAAACCTCTCGAATAGACCAAAAACAAGAGTACTAATCA
TGAAGAACAAACAAACCTCATGATCAATTACCTTATAGAAAAGCTAGGGTTTCCGTTAG
AGCTAGATCTGATGCCACTACGGTAAATGACGGATGTCAATGGAGAAAATACGGTCAGAA
AATGGCGAAAGGGAATCCATGTCTCGCGCTTATTATCGTTGCACCATGGCCGTTGGATG
TCCTGTCCGTAAACAGGTCCAACGATGCGCGGAGGATACAACCTATCTTGACAACAACGTA
CGAAGGAAACCATTAACCATCCTCTTCCCCCGTCAGCCACAGCCATGGCTGCAACCACTC
CGCCGAGCAGCCATGCTCTTATCAGGCTCCTCTCCAGCAACCTCCACCAAACACTCTC
TAGCCCCCTCCGCCACGTCATCATCATCTTCTACCATAACTTCCCATACACCTCCACAAT
CGCAACACTCTCTGCCTCAGCTCCTTCCCCACCATAACCTTAGACCTCACCACCCACC
TCGACCGCTACAACCGCCACCGCAGTTTCTAAGCCAGTATGGTCCCGCCGCGTTTACC
AAACGCTAATCAAATTAGGTCTATGAATAATAAACCAGCAGTTATTAATACCTAATTT
GTTTGGCCCAACAGCCCCACCGTGAAATGGTCGATTTCAGTTAGGGCTGCGATTGCGAT
GGATCGGAACCTTCACGGCGGCACCTGCGGCCGCGATCTCAAACATTATCGGAGGAGGTAA
TAACGACAACAATAATAACTGATATTAATGATAACAAGGTTGATGCAAAAAGTGGAGG
GAGTAGTAACGGAGATTGCCACAGCTTCTCAGTCTTGACCACTTTCTCTACAACTA
ATTTTACTACCATTATTATATGTTATCTTATTATATATTACACACATATTATACATTA
TGCGTATCTTAAGTTTTTTTTTGGGGGCCATTATATATGAATGATATGGAGATCACTGAG
AGAGAGAGAGAGCTATTATGGGTTTTTTTTT

>G1417 Amino Acid Sequence (domain in AA coordinates: 239-296)

MEEHIQDRREIAFLHSGEFLHGDSKDHQPNESPVRRHHESSIKEVDFFAAKSQPFDLG
HVRTTTIVGSSGFNDGLGLVNSCHGTSSNDGDDKTKTQISRLKLELERLHEENHKLKHL
DEVSESYNDLQRRVLLARQTQVEGLHHKQHEDVPQAGSSQALENRRPKDMNHETPATLTK
RRSPDDVDGRDMHRGSPKTPRIDQNKSTNHEEQNPHDQLPYRKARVSVRARS DATTVND
GCQWRKYGQKMAKNPCPRAYRCTMAVGCPVRKQVQRCAEDTTILTTTYEGNNHPLPP
SATAMAATTSAAAAMLLSGSSSSNLHQLTSSPSATSSSSFYHNFYPTSTIATLSASAPFP
TITLDLTNPPRPLQPPPQFLSQYGPAALPNANQIRSMNNNNQQLLIPNLFGPQAPPREM
VDSVRAAIAMDPNFTAALAAAIISNIIIGGNNDNNNTDINDNKVDAKSGSSNGDSPQLP
QSCTTFSTN*

>G1442 (1..1293)

ATGGGAACAAGAGCAGAACGCAAGGAAGATTTTGTGGTGGGTTTGGATTTGGTGTGTA
GAAATTCGCATAAAGACGTTATGGTGCTACCTCATCATCACTATTATCCATCATATTCA
TCACCTTCCTCTTCTTCTTGTGTTACTGTTCTGCTGGTGTAGCGATCCCATGTTCTCT
GTTTCTAGCAATCAGGCTTACACTTCTTCTCACAGTGGTATGTTACACCCGCCGTTCT
GTTTCTGCTGCTGTGACTGTAGCAGATCCTTTTTTCTCCTTGAGCTCTTCAGGGGAAATG
AGAAGAAGTATGAACGAAGATGCTGGTGCAGCTTTCAGCGAAGCTCAATGGCATGAGCTT
GAGAGGCAGAGGAATATATACAAGTACATGATGGCTTCTGTTCTGTTCTCCAGAGCTT
CTCACACCTTTCCCAAGAACCACCAATCAAACACTAACCCGGATGTAAGTGTGGCAGTG
GCGACAGAGGCTCAFTGCAGCTGGGGATTGCTTCAAGCGCAAGCAATAACACGGCTGAT
CTGGAGCCATGGAGGTGCAAGAGAACAGATGGGAAGAAATGGAGGTGCTCTAGAAACGTG
ATTCTGTATCAGAAATACTGTGAGAGACACACACAAGAGCCGTCCTCGTTCAAGAAAG
CATGTGGAATCATCTACCAATCATCTCACCACAATGACATTTCGTACGGCTAAGAATGAT
ACTAGCCAGCTTGTGAGAACTTATCCTCAGTTTTTACGGACAACCTATAAGCCAGATCCCT
GTGCTTTCTACTCTTCCGTCTGCCTCCTCTCCATATGATCACCACAGAGGACTGAGGTGG
TTTACGAAAGAAGATGATGCCATTGGAACTTAAACCCGGAGACTCAAGAAGCTGTCCAG
CTGAAAGTTGGATCAAGCAGAGAGCTCAAACGGGGATTTCGATTATGATCTGAATTCAGG

CAGAAAGAGCCAATAGTAGACCAGAGCTTTGGAGCATTGCAGGGTCTATTAAGTCTAAAC
CAGACACCACAACATAACCAAGAAACAAGACAGTTTGTGTAGAGGAAAGCAAGATGAA
GCGATGGGAAGCTCTCTGACACTCTCAATGGCTGGAGGAGGCATGGAGGAAACAGAGGGA
ACAAACCAGCATCAGTGGGTTAGCCATGAAGGTCCATCATGGCTCTATTCAACAACACCA
GGTGGACCATTGGCTGAAGCACTGTGTCTCGGTGTCTCCAACAACCCAAGTTCTAGTACT
ACTACTAGTAGCTGCAGCAGAAGCTCAAGCTAA

>G1442 Amino Acid Sequence (domain in AA coordinates: 172-223)
MGTRAERKEDFVGFGFVGVVENS HKDVMVLPHHHYPSYSSPSSSSLCYCSAGVSDPMFS
VSSNQAYTSSHSGMFTPAGSGSAAVTVADPFFSLSSSGEMRRSMNEDAGAAFSEAOWHEL
BRQRNIYKMMASVVPPELLLPFPKNHQSNTPDVTVAATGGSLQLGIASSASNNTAD
LEPWRCRRTDGKKWRCRNVIPDQKYCERHTHKSRPRSRKHVESHSQSSHNDIRTAKN
TSQVLRTYPQFYGPISQIPVLSLTPSASSPYDHRGLRWFTKEDDAIGTLNPETQEAVQ
LKVGSSRELKRGFDYDLNFRQKEPIVDQSFGALQGLLSLNQTPQHNQETRQFVVEGKQDE
AMGSSSLTSLMAGGMEETEGTNQHWVSHEGPSWLSTTPGGPLAEALCLGVSNPNSSST
TTSSCSRSSS*

>G1454 (86..1180)
CTAGTAGTGATGATATGATCGCTTCTTCTCCTACAATCTCAGAAAACCTCCGATCACGGTT
TTAGATATCTTCTACAACGGATACAATGGAGAGCACCGATTCTTCCGGTGGTCCACCACC
GCCACAACCTAACCTTCTCCTCCAGGCTTCCGGTTTACCCTACCGACGAAGAGCTTGTGT
TCACTACCTCAAACGCAAAGCAGCCTCTGCTCCTTACCTGTCCGCATCATCGCCGAAGT
CGATCTCTATAAATTTGATCCATGGGAACCTCCCGCTAAAGCATCGTTTGGAGAACAGA
ATGGTACTTCTTTAGTCCACGAGATCGGAAGTATCCAAACGGAGCAAGACCAACAGAGC
GGCGACTTCAGGTTATTTGGAAAGCGACCGGTACAGATAAACCGGTACTTGCTTCCGACGG
TAACCAAAAGGTGGGCGTGAAGAAGGCACTAGTCTTCTACAGTGGTAAACCACCAAAAGG
CGTTAAAGTGATTGGATCATGCATGAGTATCGTCTCATCGAAAAACCAACAAATCG
ACCTCCTGGCTGTGATTTCCGGCAACAAAAAACTCACTCAGACTTGATGATTGGGTGTT
ATGTAGAATCTACAAGAAGAACAACGCAAGTCGACATGTTGATAACGATAAGGATCATGA
TATGATCGATTACATTTTCAGGAAGATTCTCCTCGTCTTTATCAATGGCGGCTGCTTCTAC
AGGACTTCACCAACATCATCATAATGTCTCAAGATCAATGAATTTCTTCCCTGGCAAATT
CTCCGGTGGTGGTTACGGGATTTTCTCTGACGGTGTAAACAGAGTATATACGACGGCGG
TGGCATGATCAACAATATTGGTACTGACTCAGTAGATCACGACAATAACGCTGACGTCGT
TGGTTTAAATCATGCTTCGTCGTCAGGTCCTATGATGATGGCGAATTTGAAACGAACTCT
CCCGGTGCCGTATTGGCCTGTAGCAGATGAGGAGCAAGATGCATCTCCGAGCAAACGGTT
TCACGGTGTAGGAGGAGGAGGAGGAGATTGTTTCAACATGTCTTCTCCATGATGGAAGA
GACTCCACCATTGATGCAACAACAAGCTGGTGTGTTAGGAGATGGATTATTCAGAACGAC
ATCGTACCAATTACCCGGTTTAAATTTGGTACTCTTCTTAATCAATGTGTTTCGCCGCGG
GTGTGAAGAATTTCCCGTGACAGTGAAGATTTTTCCTGATTGGTGGGGTCATTGTCAT
GCATTATATAATTTGAGATTTGTGTATATGTTTTGGGTTAATTAATTGGTCACAGGGGC
>G1454 Amino Acid Sequence (conserved domain in AA coordinates: 9-178)
MESTDSSSGPPPPQPNLPPGFRFHPDEELVVHYLKRKAASAPLPVAIIAEVDLYKFDPW
ELPAKASFGEQEWYFFSPDRKYPNGARPNRAATSGYWKATGTDKPVLASDGNQKVGKK
ALVFYSGKPPKGVKSDWIMHEYRLIENKPNRPPGCDGFKNKNLRLDDWVLCRIYKNN
ASRHVDNDKDHMDIFYFRKI PPSLSMAAATSTGLHQHHNVSRSMNFFPGKFSGGGYGIF
SDGGNTSIYDGGGMINNIGTDSVDHNDADVGLNHASSSGPMMANLKRTPVPYWPVA
DEEQDASPSKRFHVGVGGGGDCSNMSSSMMEETPPLMQQGGVLGDGLFRTTSYQLPGLN
WYSS*

>G1459 (1..1272)
ATGATGAAAGGTCTGATTGGGTATAGATTTAGTCCGACGGGAGAGGAAGTGATCAACCAT
TACCTAAAGAACAAACTTCTGGGTAAGTATTGGCTCGTTGATGAAGCTATTAGCGAGATC
AACATCTTGAGTCACAAACCCAGCAAGGATTTGCCTAAGTTAGCTAGGATCCAATCGGAA
GATCTTGAATGGTATTCTTCTCTCCGATTGAGTACACGAACCCGAATAAGATGAAAATG
AAGAGGACGACAGGTTCTGGGTTTTGAAACCTACTGGTGTGATCGGGAAATTAGGGAT
AAAAGAGGAAATGGTGTGTGATAGGGATTAAGAAGACGCTTGTGTACCATGAAGGTAAG
AGTCCTCATGGAGTTAGAACTCCTTGGGTTATGCACGAGTATCACATCACTTGCTTGCT
CATCATAAGAGGAAATATGTTGTCTGCCAAGTAAAGTATAAGGGTGAAGCTGCAGAAATT
TCATATGAGCCCAAGTCCCTCTTTGGTATCCGATTGCGATACCGTCATAGCGATTACCGGA
GAACCGGAACCTGAGCTTCAGGTTGAGCAGCCAGGTAAAGAAAATCTCTTGGGTATGTCT

GTAGATGATTTGATAGAACCAATGAACCAACAAGAGGAGCCACAAGGTCCTCACTTAGCT
CCGAATGATGATGAGTTTATACGTGGATTGAGGCATGTTGATCGAGGGACGGTTGAATAT
TTGTTTGCCAATGAAGAAAACATGGATGGTTTGTCTATGAATGACTTGAGAAATCCCAATG
ATCGTCCAACAAGAGGATCTCTCTGAGTGGGAGGGATTTAACGCAGACACCTTTTTCAGC
GACAACAACAATAACTATAACCTTAACGTGCATCATCACTAACGCCTTACGGCGATGGC
TATTTGAATGCATTTTCGGGTTATAACGAAGGGAATCCTCCCGATCACGAATTAGTGATG
CAAGAGAACCGCAACGATCACATGCCAAGGAAACCTGTGACAGGGACCATTGATTATAGC
AGCGATAGTGGCAGTGATGCTGGATCCATATCTACAACGGTGAAACAAGAAATCCCAAGA
GCTGTTGATGCACCCATGAACAATGAGTCATCTTTGGTGAAAACAGAGAAGAAAGGCTTG
TTTATTGTAGAGGACGCAATGGAGAGAAACCGCAAGAAACACGATTTATCTATCTCATG
AAGATGATCATAGGCAACATCATATCGGTTTTACTACCCGTCAAAAGATTGATCCCGGTG
AAGAAGTTATGA

>G1459 Amino Acid Sequence (conserved domain in AA coordinates:10-152)

MMKGLIGYRFSPTGEEVINHYLKNKLLGKYWLVEAISEINILSHKPSKDLPLKARIQSE
DLEWYFFSPIEYTNPNKMKMKRTTSGSFWKPTGVDRIRDKRNGVVGIGIKKTLVYHEGK
SPHGVRTPWVMHEYHITCLPHHKRKYVVCQVKYGEAAEISYEPSPLVSDSHTVIAITG
EPEPELQVEQPGKENLLGMSVDDLI EPMNQEEPPQGPLAPNDDEFIRGLRHVDRGTVEY
LFANEENMDGLSMNDLRIPMIVQQEDLSEWEGFNADTFFSDNNNNYNLNVHQLTYPYDGG
YLNAFSGYNEGNPPDELVMQENRNDHMPRKPVGTIDYSSDSGSDAGSISTTVKQEI PR
AVDAPMNNESLSVKTEKKGLFIVEDAMERNRKKPRFIYLMKMIIGNIISVLLPVKRLIPV
KKL*

>G1460 (87..995)

CGTCGACCTTCACTCAAACCTTAATCCCGGGAACCCGGGAATTTTGATCATTTTGTCTTCT
TTTCGATCTGTTTCTATTTTAAAAAGATGATGAAAGATCCGACTGGGTATAGATTAGTC
CGACGGGAGAGGAAGTGATAAACCATTAACCTAAAGAACAAAATTTGCGGTAAGACTTGGC
TCGTTGATGAAGCCATTAGCGAGATCAACATCTTGAATCACAACCCAGCAAGGATTGTC
CTAAGTTAGCTAGGATCCAATCGGAAGATCTTGAGTGGTACTTTTCTCTCCGATTGAGT
ACACGAACCCGAATAAGATGAAAATGAAGAGGACGACAGGTTCTGGGTTTTGGAAACCTA
GTGGTGTGATCGGAAAATTAGGGATAAAAAGAGGAAATGGTGTGTGATAGGGATTAAGA
AGACGCTTGTGTACCATGAAGGTAAGAGTCCCTCATGGAGTTAGAACTCCTTGGGTTATGC
ACGAGTATCACATCACTTGCTTGCCCTCATCATAAGAGGAAATATGTTGTCTGCCAAGTAA
AGTATAAGGGTGAAGCTGCAGAAATTTTCATATGAGCCAAGTCCCTCTTGGTATCCGATT
CGCATACCGTCATAGCGATTAACGGAGAACCGGAACCTGAGCTTCAGGTTGAGCAGCCAG
GTAAAGAAAATCTCTTGGGTATGTCTGTAGATGATTTGATAGAACCAATGAACCAACAAG
AGGAGCCACAAGGTCCTCACTTAGCTCCGAATGATGATGAGTTTATACGTGGATTGAGAC
ATGTTGATCGAGAGCCGGTTGAATATTTGTTTGCCAATGAAGAAAACATGGATGGTTTGT
CTATTATGAATGACTTGACAATCCCAATGATCGCCCAACAAGAGGATCTCATTCTCTCTG
AGTGGGAGGGATTATCGCAGCCACCTTTTTCAGCGACAACAACAATAACAATAACCTTA
ACGTGCATCAACTAACGCTTTTCTTACC GGATGATTATCAGAATGCATTTTGGGTTACA
ACGGAGCGNCCGCT

>G1460 Amino Acid Sequence (domain in AA coordinates: TBD)

MMKDPTGYRFSPTGEEVINHYLKNKILGKTWLVDEAISEINILNHKPSKDLPLKARIQSE
DLEWYFFSPIEYTNPNKMKMKRTTSGSFWKPSGVDRKIRDKRNGVVGIGIKKTLVYHEGK
SPHGVRTPWVMHEYHITCLPHHKRKYVVCQVKYGEAAEISYEPSPLVSDSHTVIAING
EPEPELQVEQPGKENLLGMSVDDLI EPMNQEEPPQGPLAPNDDEFIRGLRHVDREPVEY
LFANEENMDGLSIMNDLTIPMIAQQEDLILSEWEGFIAATFFSDNNNNNNLNVHQLTSL
PG*

>G147 (37..672)

AAATCATCAGATAGAAGGAAATATTCTGATTGAGAGATGGCTCGTGGAAGATTGAGCTT
AAGAGGATTGAGAACECGGTTACAGACAAGTGACTTTTTCAGAGGAGAACTGGTCTT
CTCAAGAAGGCTAAGGAGCTCTCTGTGCTCTGTGATGCCGAGATCGGTGTTGTGATCTT
TCTCCTCAGGGCAAGCTCTTTGAGCTCGCTACTAAAGGAACAATGGAGGGAATGATTGAT
AAGTACATGAAGTGTAAGTGGTGGTGGTGGTCTGTTCTTCTGCTACTTTTACTGCTCAA
GAACAACCTTCAACCACCAATCTTGATCCGAAAGATGAGATCAACGTGCTTAAGCAAGAG
ATTGAGATGCTTCAGAAAGGATAAGCTATATGTTTGGAGGAGGAGATGGGGCTATGAAT
CTTGAAGAACTTCTTTGCTTGAGAAGCATCTTGAGTATTGGATTTCTCAGATTTCGCTCT
GCTAAGATGGATGTTATGCTTCAAGAAATTCAGTCATTGAGGAACAAGGAAGGAGTCCTC

AAAAACACCAACAAGTATCTCCTCGACAAGATAGAGGAAAAACAACATAGCATATTAGAT
GCTAACTTCGCAGTCATGGAGACAAACTATTCCTATCCGCTAACAAATGCCAAGTGAAATA
TTTCAGTTCTAGACCATAGGGTATTTGAAGACTATGTCTCACGAATTTAAATAACCTTGG
TAAGTATAATATAGTGTGTAAATCACACATAATTAAATAAAGCCTGTGGAACCTCGC
TAGGCAGTTGAAAATCTATCCGTATGTTTTATCCTCTTGTTTTACATTTGTTGGTGTGAA
GATGAAATGACTGCAAGTGTGGTGTGTACTTATAACTCTTCTACTTTCTATCTATGTTT
TGAATTTATGGATT

>G147 Amino Acid Sequence (domain in AA coordinates: 2-57)
MARGKIQLKRIENPVHRQVTFCKRRTGLLKKAKELSVLCDAEIGVVIFSPQGLFELATK
GTMEGMIDKYMKTGGGRGSSATFTAQEQLOPPNLDPKDEINVVKQEIEMLQKGISYMF
GGGDGAMNLEELLLEKHLEYWISQIRSAKMDVMLQEIQSLRNKEGVLKNTNKLKLLDKIE
ENNNSILDANFAVMEITNYSYPLTMPSEIFQF*

>G1471 (1..735)

ATGGAGAACCAATCTATGTCTTCATCAAGCTCCTCCACACACAAACATGATCAAAAACCTC
AAAAGTTCGGTGTGGCCATGGAGGTCTTGGAGGAGAAGGAGACAGTGAACAATCCGCCC
CAGTATTATAATAAGATCTACATCTGTTACTTGTGCAAGAGAGCGTTCCCAACCCCTCAT
GCCCTTGGCGGTACGGAACCCACCAAGGAGGACCGAGAATTGGAGAGGCAACAGATC
GAGTCAAGGCTTTCTAACAAAGACAAGTCTAACTTGCTCTTGGTGGGTCTTCACAAGAT
GTTTTATCAAATGATAATCACCTTGGACTCTCTCTTGGTCCATTGAAGTCCATAGAAGGT
AGCAGCAGCAGCAACAACGTTAACCATTGCTTAATGTTGGAGTCCCTAGAGGAACCACA
GATATGAACATGAACAACATATAGCTCACATGCTTTATCAACTGATGATATTAATCTTGAT
CTTACTCTTGGTCCATCTAAGTCCATAGGAGATAGCAACAATATCATTAAATAACAACACT
AATCATCTCTCGATGGGAATCTGATCATTCCTCGTTCCTCGTGTGTCTAGATACCAT
TTTGTGTGCTGGGAACCCCTTGATTCAATCTCTAGAAACATTCTCCTTCTATTACTTTT
CCTCATCTAAACATCAATCTTCTCATGATTGTTTTCTTTACAAGAGAATGGTTCGGGC
TCTAGTCACTCATAA

>G1471 Amino Acid Sequence (domain in AA coordinates: 49-70)
MENQSMSSSSSTHKHDQKLKSSVVAMEVLEEKETVNNPPQYYNKIYICYLCKRAFPTPH
ALGGHGTTHKEDRELERQQIESRLSNKDKSNLLFGGSSQDVLSDNHLGLSLGPKLSIEG
SSSSNNVNPLLNVGVPRGTTDMNMNYSHALSTDDINLDTLGPSSKISGDSNNIINNNT
NSSFDGNLIIPVRPRVSRVYHFVAGNPLDSISRNPPTSITFPHLNINLSHDSFSLQENGSG
SSHS*

>G1475 (1..645)

ATGAAGAGAACACATTTGGCAAGTTTTAGTAACAGAGACAAAACCCAAGAAGAAGAAGGA
GAAGACGGTAATGGTGACAACAGAGTCATCATGAATCACTACAAGAATTACGAAGCTGGG
CTGATCCCATGGCCTCCCAAGAATTACACTTGCAGCTTCTGCAGGAGAGAGTTTCAATCT
GCTCAAGCACTTGGAGGCCACATGAATGTTCATAGAAGAGACAGAGCAAACTCAGGCAG
ATCCCTTCTTGGCTCTTTCGAACCTCACCACCACACACCTATTGCAAACCCTAACCCCTAAT
TTTAGCTCTTCTTCTTCTTCTTCAACAACAACAGCTCATCTTGAGCCTTCCCTAACCAAC
CAGAGATCCAAAACCACTCCTTTCTTCTTCTGCCCCGTTTGATCTTTTGGACAGTACTACT
AGCTATGGAGGTTTGATGATGGACAGAGAGAAGAACAAGAGCAATGTATGTAGCAGAGAG
ATCAAGAAAAGTGCCATCGATGCATGTCATTAGTAAGATGTGAGATAAGCCGTGGGGAT
CTGATGAATAAGAAAGATGATCAAGTCATGGGGTTGGAGCTTGGGATGAGTTTGAGGAAT
CCCAACCAAGTCTTGGATTGAGCTTGGAGCTTGGAGCTAGGCTACCTCTAA

>G1475 Amino Acid Sequence (domain in AA coordinates: 51-73)
MKRTHLASFSNRDKTQEEEGEDGNGDNRVIMNHYKNYEAGLIPWPPKNYTCSFCRREFRS
AQALGGHMNVHRRDRAKLRQIPSWLFEPHHHTPIANPNPNFSSSSSTTTAHLEPSLTN
QRSKTTFPFSARFDLDSTTSYGGMLMDREKNKSNVCSREIKKSAIDACHSVRCEISRGD
LMNKKDDQVMGLELGLMLRNPNQVLDLELRLGYL*

>G1477 (1..606)

ATGTTGTCTCGGACTCGAATTACGCTAGTGATATTAGCGACGATGCCTCCGCCACCGGA
TCGATAGAGAATCCTATATACAAATGCAAGTATTGTCTAGGAAGTTGATAAAACACAA
GCATTAGGTGGTCATCAAAATGCACACAGAAAGGAGAGAGAGGTGAAAAACAACAAAA
GCATTTTGGCGCATTTGAACGACCAGAACAGATCTTTACGCGTACTCGTATTCTGTAT
CATCATTCATTTCTTAACCAATACGCACCTCCACCGGGATTGTAACAGCCTCAGTACAAA
GTTGATAGATCATACAAGATGTCATGGTCTACAACCAATATGTGGGATCCTCAAGCTCT
AGCTTTGCAGGACTACAAAGTGACCCAAGTCAAGGAATGAACCAGGATTGGACCTTTACC

GGGATCCCATTCTACCCCAATCTCAACCTCAACCACTATCGTCACCAATATGTTTGGAT
CTTTGCCTTGGCATTTGGTAGCTCCCAAACCCAACCACAACCTCAAGAACCAATGATGCA
ACAGAAGAGATGGATGCTGAGAAAGAAAATGATGGTTCTTCCCTTTCTCTCACTCAAA
CTGTGA

>G1477 Amino Acid Sequence (domain in AA coordinates: 29-48)
MLSSDSNYASDISDDASATGSIENPIYKCKYCPRKFDKQALGGHQNHRKEREVEKQKQ
AFLAHLNRPEPDLAYSYSYHHSFPNQYALPPGFEPQYKVDRSYKMSMVYNQYVGSSES
SFAGLQSDPSQGMNQDWTFTGIPFLPQSQPQLSSPICLDLCLGIGSSQTQPQPQEPNDA
TEEMDAEKENDGSSLSLSLKL*

>G1487 (1..1020)
ATGGAACAAGCCGCGTTGAAGAGCAGCGTCAGGAAAGAGATGGCTCTCAAAACGACTTCT
CCGGTTTACGAAGAGTTTCTTGCCGTCACCACCGCTCAAAATGGCTTTTCCGTCGACGAT
TTCTCTGTAGACGACTTGCTTGACTTGTCAAACGATGACGTTTTTGGCCGACGAAGAACT
GACCTCAAGGCTCAACATGAGATGGTCCGTGTTTCTCTGAGGAACCCACGACGACGGA
GACGCTCTTCGCCGAGCAGCGATTCTCCGGCTGTGACGACTTTGGTTCTCTCCCTACA
AGCGAATCTCTCTTCGCCGCGATGATTTAGCGAACCTTGAGTGGCTCTCTCATTTCTGTG
GAGGACTCTCTTCACGGAATATTCGGGTCCAAACCTCACCAGAACCCCGACTGAGAAACCG
GCGTGGTTAACGGGTGACCGGAAACATCCTGTGACTGCAGTCACGGAAGAGACCTGTTTC
AAATCCCCTGTTCGGCTAAAGCCCGTAGCAAACGTAACCGCAATGGCCTCAAGGTCTGG
TCGCTTGGTTCTGTCCTCTCCGGTCTCTCGTCCGGTTCGACCTCTCTCTCTCTCT
TCGGGTCTTCCAGCCCGTGGTTCTCCGGCGCTGAGCTGCTCGAGCCTGTGGTCACGTCA
GAGAGGCCACCGTTTCCCAAGAAGCATAAGAAAAGGTGAGCCGAGTCTGTTTTCTCCGGT
GAGCTGCAGCAGCTGCAACCTCAGCGAAAGTGCAGCCACTGCGGCGTTTCAAGAACTCCG
CAGTGGAGAGCCCGGCAATGGGAGCCAAAGACCTGTGCAATGCGTGCCTGTCTCCGGTAC
AAGTCCGGTAGGTTGCTACCGGAATACAGACCCGCTTGTAGCCCGACATTCTCGAGTGAG
CTGCACTCGAACCAACCCGGAAGTCAAGAGATGAGGCGGAAGAAGGAGCAACCACT
GACAACGAAACCGGTTTAAACAGCTGGTTTCACTCCCAAGCTGTACCAAGTTTTTGA
>G1487 Amino Acid Sequence (domain in AA coordinates: 251-276).
MEQAALKSSVRKEMALKTTSPVYEEFLAVTTAQNGFSVDDFSVDDLLDLSNDDVFADDET
DLKAQHEMVRVSEEPNDDGDLRRSSDFSGCDDFGLPTSELPLPADDLANLEWLSHFV
EDSFTEYSGPNLTGTPTKPAWLTGDRKHPVTAVTEETCFKSPVPAKARSKNRNGLKVN
SLGSSSSSGPSSSGSTSSSSSGPSSPWFSGAELLEPVVTSEPPFPKHKHRSAESVFSG
ELQQLQPKRCSHCGVQKTPQWRAGPMGAKTLCNACGVRYKSGRLLPEYRPACSPFSSSE
LHSNHRKVIEMRRKKEPTSDNETGLNQLVQSPQAVPSF*

>G1492 (149..919)
AATCCCAACCCACACACCTCTCAAATCCTCCTCTCCTCGTTTCTCTTTCTCTCTCTTCA
CAGAACCAAAACATATCAAACCTTTTTTCTCTTGGGTTTAAAGTAAAAATCGAATCTTTG
TCTCGGTTTTTGGGTTCTTGAACGATATGGGTAAGTCTAGTGGTAGAAATGGTAACGG
AAGCTTTAACGGCAATAAATTTACCGGAGTTAGACCTTACGTACGGTCTCCAGTTCCACG
GCTTAGATGGACGCCGATCTTACCGTTGTTTCGTTACGCCGTCGAGATTCTCGGTGG
TCAACACCGAGCAACACCAAACTTGTTCTTAAGATGATGGATGTGAAGGGACTTACCAT
TTCACATGTCAAAAGCCACCTTCAAGATGTATAGAGGAGTTCAAAGCTCACTTTGGAGAA
ACCAGAAGAAAGCTCATCATCTTCAATAAGAAGAAGACAGACAGTGAAGAAGATTATTA
TCTTCATGACAACTTGTCTTTACACACAAGGAATGATTGTCTTTTGGGTTTCACTCTTT
TCTCTTTTCTTACATTCTTCAATTAGAGGAGGAGGAGGAGGAAGAACAAGAGCAGCA
GACTTCAGAGTCTGGTGGTTATGATGATGATGCTGACTTCTTACATCAAGAAGATGAA
CGATACGACGACGTTTTTGTACATCATTTCCCAAGGGAACAGAGGAGTGGCGGGAACA
AGAACACGAAGAAGAAGAAGATTGTCTGTTGTCTCTGTCGTTAAATCATCATATTG
GAGAAGCAATGATCATCGGTGGTGAGCGAAACGAGTGAAGCAGCAGTCTCGACTTGTT
AGCACCATTCTGATCCAAAGATTGCTTTGGTTCTTCAAAGATTGATCTTAATCTGTCAAT
TTCTCTCTCGGTAGCTAAATAAGTTATGCAAGATTAGGTTTCAAGAACTATTCCGAT
GTGTTTTTGAAGTATGATATTGAATGTTAGTAGAGAACTAGAAATGAAGTTTAGAT
AAATTATCAACGACGCTTTTGTGCTGCTTTGAACGGAATAAACA

>G1492 Amino Acid Sequence (domain in AA coordinates: 34-83)
MGKSSGRNGNSFNNGKFKHGVPRPVPRPRLRWPDLHRCFVHAVEILGGQHRATPKLV
LKMMDVKGLTISHVKSHLQMYRGSKLTLEKPESSSSSIRRRQDSEEDYVLDNLSLHT
RNDCLLGFHSFPLSSHSSFRGGGGGRTKEQQTSESAGYDDADFLHIKMDNTTFLSHH

FPKGTEEWREQEHEEEEEEDLSLSLSLNHHHWSNGSSVVSETSEAAVSTCSAPFVSKDCF
GSSKIDLNLISISLLGS*

>G1531 (1..666)

ATGTGTGAGTCAAGCAACAAAGTCAGAGTATCGCCATACCCGCTTCGGTCTTCGAGGACC
GACAAACACAAGGCGTCAGAGTCGCCTATTGAGACAGGTTGGGAGGATGTGCGTGGATGT
CATCCTTACATGTGCGATACGAGTGTTTCGTCACTCCAATTGTTTCAAGCAGTTCGCGAGA
AAAACCATAAAAAAGCGCCTATACCCCAAGACCTTACATTGTCCTCTCTGTAGAGGTGAA
GTATCCGAGACGACAAAGGTGACGAGCACTGCAAGAAGATTATGAATGCTAAACCGAGG
TCTTGCTCCGTAGAGGATTGCAAATTCTCTGGGACGTTTTCTCAGCTTACTAAGCACTTG
AAAACAGCATCGCGGTATTGTGCCACCAAGGTGATCCACTGAGACAACAGAGATGG
GAAATGATGGAGAGACATTCTGAATACGTTGAACTCATGACTGCAGCTGGGATTTCGCGT
ATGGCTGAGGTGATGCAACAACAGCTTCCCCAGGATCAGAAATCATCTCATGTGTTCAA
GTGACCGTTAATGGAACCATATGGAATCTAATTGATCCGAGTCAGGGAAGGAATGGATTA
GGCATACCAACTATAGCGCAATGCAGTTTGTACCATTAAAGCATAAATCACAGTAGAACT
CTGTGA

>G1531 Amino Acid Sequence (domain in AA coordinates: 41-77)
MCESSNKVRVSPYPLRSSRTDKHKASESPIETGWEDVRGCHPYMCDTSVRHSNCFKQFRR
KTIKKRLYPKTLHCLCRGEVSETTKVTSTARRFMNAKPRSCSVEDCKFSGTFSQLTKHL
KTEHRGIVPPKVDPLRQQRWEMMERHSEYVELMTAAGISRMAEVMQQQLPQDQNHPIVFO
VTVNGTIWNLDIPSQGRNGLGITNYSAMQFVPLSINHSTRLL*

>G1540 (122..997)

atctctttactaccagcaagttgttttcttgtaacttcaaacttctctttctctgttc
ctctctaagtcttgatcttatttaccgttaactttgtgaacaaaagtcgaatcaaacaca
catggagccgccacagcatcagcatcatcatcaagccgaccaagaaagcggcaacaa
caacaacaagtcgggctctgggtgttacacgtgtcgccagaccagcacgaggtggacacc
gacgacggagcaaatcaaaatcctcaaagaactttactacaacaatgcaatccggtcacc
aacagccgatcagatccagaagatcactgcaaggctgagacagttcggaaagattgaggg
caagaacgtcttttactggttccagaaccataaggctcgtgagcgtcagaagaagagatt
caacgggaacaaacatgaccacaccatcttcacccaactcgggttatgatggcggctaa
cgatcattatcatcctctacttaccatcatcacgggtgttcccatgagagacctgctaa
ttccgtcaacgttaaaacttaaccaagaccatcatctctatcatcataacaagccatatcc
cagcttcaataacgggaatttaaatcatgcaagctcaggtactgaatgtggtgtgttaa
tgcttctaattggctacatgagtagccatgtctatggatctatggaacaagactgttctat
gaattacaacaacgttaggtggaggtgggcaaacatggatcatcattactcatctgcacc
ttacaacttcttcgatagagcaaagcctctgtttggtctagaaggtcatcaagacgaaga
agaatgtggtggcgatgcttatctggaacatcgacgtacgcttccctcttccctatgca
cgggtgaagatcacatcaacgggtggttagtggtgccatctggaagtatggccaatcggaagt
tcgcccttgcgcttctcttgagctacgtctgaactagctcttacgcgggtgtcgctcggg
attaaagctcttccctctctctctctctctctctgactcgtatgttcacaactatgcttcgc
tagtgattaatgatgcagttgttatattagtagttaactagttatctctcggttatgtgta
atttgtaattactagctaagtatcgtctaggtttaattgtaattgacaaccggttatctc
tatgatgaataagttaaatttatatat

>G1540 Amino Acid Sequence (domain in AA coordinates: 35-98)
MEPPQHQQHHHQADQESGNNNKGSGGYTCRQTSTRWTPTEQIKILKELYNNNAIRSP
TADQIQKIKITARLRQFGKIEGKNVYWFQNHKARERQKRFGNTNMTTPSSSPNSVMMAAN
DHYHPLLHHHHGVPMQRPANSVNVKLNQDHHLYHHNKPYPSPFNNGNLNHASSGTECGVVN
ASNGYMSSHVYGSMEQDCSMNYNVGGGWANMDHHYSSAPYNFFDRAKPLFGLEGHQDEE
ECGGDAYLEHRRTLPLFPMHGEDHINGGSGAIWKYQSEVRPCASLELRNLN*

>G1544 (1..2178)

ATGTCTCAGTCAAACATGGTACCAGTGGCTAACCAACGGAGACAACAACAGACAACGAA
AACAAACAACAACAACAACAATGGTGGAACTGACAACACTAATGCTGGAAATGATTCT
GGAGATCAAGATTTTCGACAGTGGGAATACCTCAAGTGGCAATCATGGAGAAGGGTTGGGA
AACAAATCAAGCTCCTCGTCATAAGAAGAAAAATACAATCGTCACACCCAACCTCAGATT
TCGGAGATGGAAGCTTTCTTCAGAGAGTGTCCTCACCCAGATGACAAACAAAGGTACGAC
CTTAGCGCTCAATTGGGATTGGACCCCTGTTAGATCAAATTCGTTCCAGAACAAACGC
ACTCAAAACAAGAATCAACAAGAACGCTTTGAGAACTCAGAACTTCGGAATCTGAACAAC
CACCTTAGGTCTGAAAATCAGCGGTTACGAGAAGCTATTCATCAAGCCTTATGCCCTAAG

TGTGGAGGCCAACTGCAATTGGCGAAATGACCTTCGAAGAGCACCATCTTCGCATCCTC
AACGCTCGTTTGACTGAAGAGATCAAGCAACTTTCGCTGACAGCGGAAAAGATATCAAGG
CTTACGGGGATACCAGTAAGGAGCCATCCCCGTGTGTCTCCTCTAATCCTCCTCCAAAT
TTCGAGTTCGGGATGGGATCTAAGGGAAATGTCGGAACCACTCGAGGGAAACCACTGGA
CCTGCAGATGCTAATACCAAGCCGATCATCATGGAGTTGGCATTGGAGCCATGGAGGAG
CTCTTGGTGTATGGCTCAAGTGGCTGAACCACTGTGGATGGGAGGATTTAATGGCACTAGC
TTAGCTTTGAACCTTGGATGAATACGAAAAGACGTTTCGCACGGGTCTCGGTCCCTAGACTT
GGCGGGTTTCGAACCGAGGCATCCAGGGAACTGCACTCGTGGAATGTGTCTCTACTGGC
ATTGTTGAAATGCTCATGCAAGAGAATCTGTGGTCAACAATGTTGCGCGAATTGTGGT
AGAGCCAGGACTCATGAACAGATAATGGCTGATGCTGCTGGAACCTTCAATGGAAATCTC
CAAATAATGAGTGCTGAGTACCAAGTGCTTTCCCCGCTAGTCACAACCCGCGAAAGCTAC
TTCGTCCGCTACTGTAAGCAACAAGGAGAGGGTTGTGGGCGGTGGTGCATATTTCCATC
GACCATCTCCTCCCAAACATCAACCTAAATGTGCGCCGCGACCTCTGGATGTCTGATT
CAAGAAATGCATAGTGGTTACTCCAAGGTTACATGGGTGGAACATGTGGAAGTAGATGAT
GCAGGAAGTTACAGCATCTTTGAGAAATTAATCTGTACTGGTCAAGCTTTTGCTGCTAAC
CGCTGGGTTGGTACATTGGTACGCCAGTGTGAGCGGATATCTAGCATCTTGTGCACAGAT
TTTCAATCTGTGCGATTCCGGTGATCACATAACGCTAACTAACCATGGAAAGATGAGCATG
CTGAAGATAGCTGAGCGGATTGCGAGAACCTTCTTTGCTGGAATGACCAATGCGACGGGG
TCTACAATATTTTCTGGTGTGTAAGGAGAAGATATCAGAGTGATGACAATGAAGAGCGTG
AATGATCCAGGAAAGCCTCCCGGTGTATTATTTGTGCAGCCACTTCTTTTGGCTTCTCT
GCTCCTCCTAACACTGTCTTTGACTTCTCTCAGAGAGGCTACTCACCAGACAAATTGGGAT
GTTCTCTGCAACGGAGAGATGATGCACAAGATAGCAGAGATTACGAATGGGATAGACAAA
AGGAAGTGTGCAAGTTTACTCCGGCATGGACACACTAGCAAGAGCAAGATGATGATAGTT
CAAGAGACTTCTACTGACCCCAACAGCTTCATTTGTGCTTTATGCGCCTGTTGATATGACA
TCAATGGATATTACTCTCCATGGAGGTGGTGTATCCTGACTTTGTGGTGATCCTGCCTTCT
GGTTTTGCTATTTTCCAGATGGTACGGGTAAGCCTGGAGGAAAAGAGGAGGATCACTT
TTGACCATTCTCTTCCAAATGCTGGTTGAGTCAGGTCCTGAGGCTAGGCTGAGTGTAGC
TCTGTTGCAACTACTGAGAATCTGATTCTGTAACCGTGCGGAGGATCAAAGATTTGTTT
CCTTGTGCACTGCTTGA

>G1544 Amino Acid Sequence (domain in AA coordinates: 64-124)

MSQSNMVPVANNGDNNNDNENNNNNNNNGGTDNTNAGNDSGDQDFDSGNTSSGNHGEGLG
NNQAPRHKKKKYNRHTQLQISEMEAFFRECPHPDDKQRYDLAQLGLDPVQIKFWFQNK
TQNKQQRERFENSELRNLNHLRSENQRLREAIHQALCPKGGQTAIGEMTFEEHHLRLIL
NARLTEEIKQLSVTAEKISRLTGIPVRSHPVSPNPNNPFEFMGSGKGNVGNHSRETTG
PADANTKPIIMELAFGAMEELLVMAQVAEPLWMGGFNGTSLALNLDEYEKTFRTGLGPRL
GGFRTEASRETALVAMCPTGIVEMLMQENLWSTMFAGIVGRARTHEQIMADAAGNFNGNL
QIMSHEYQLVSLPLVTRESYFVRYCKQQGEGWLAVVDISIDHLLPNINLKCRRRPSGCLI
QEMSHGYSKVTWVEHVEVDDAGSYSIFEKLICTGQAFANRWGTLVRQCERISSILSTD
FQSVDSGDHITLTNHGKMSMLKIAERIAITFFAGMTNATGSTIFSGVEGEDIRVMTMKS
NDPGKPPGVIIICAATSFWLPAPPNTVFDLREATHRNWDVLCNGEMMHKIAEITNGIDK
RNCASLLRHGHTSKSKMMIVQETSTDPTASFVLYAPVDMTSMIDITLHGGGDPDFVILPS
GFAIFPDGTGKPGGKEGGSLLTISFQMLVESGPEARLSVSSVATTENLIRTTVRRIKDLF
PCQTA*

>G156 (39..755)

AGGAAGAGGGAGCCACTCATAAGAGGAAGAAGAGAGAGATGGGTAGAGGGAAAGATAGAGA
TAAAGAAGATAGAGAATCAGACGGCGAGGCAAGTGACCTTCTCCAAGAGAAGAACTGGTC
TTATAAAGAAGACTCGTGAGCTCTCTATTCTCTGTGACGCTCACATCGGTCTCATCGTCT
TCTCAGCCACCGGAAAGCTTTCGAGTTCTGCTCCGAACAGAACAGGATGCCTCAACTCA
TTGACCGATACTTGATACCAACCGEATTGCGACTTCCTGATCATCATGACGACAGGAGC
AATTGCACCATGAGATGGAACCTACTAAGAAGAGAGACATGTAACCTTGAGCTTCGTCTGC
GTCCATTCCATGGACATGACTTAGCCTCCATTCTCTTAATGAGCTTGACGGACTCGAGA
GACAGCTAGAACATTCTGTCTCAAGTCCGTGAGCGTAAGAGGAGGATGCTAGAGAAG
ATAACAACAACATGTACCGTTGGCTTTCATGAGCATCGTGCAGCGATGGAGTTTCAACAAG
CTGGGATAGATACCAAACAGGGGAGTATCAACAGTTTATAGAGCAGCTTCAGTGCTATA
AACCAGGGGAGTATCAGCAGTTTCTAGAGCAGCAGCAACAACAACCAACAGCGTTCTTC
AGCTTGCTACACTTCTTCTGAGATTGATCCTACTTACAATCTCCAGCTTGCTCAGCCTA
ATCTTCAAAACGATCCAACGGCCCAAGATGATTAATACAATTCTCAATAGATATCTACTC

[illegible]

TGTTACTACTACGAGTTCATGCCTCTGAAGAACTGA
>G1587 Amino Acid Sequence (conserved domain in AA coordinates: 61-121)
MGYISNNNLINYLPLSTTQPPLLLTHCDINGNDHHQLITASSGEHDIDERKNNIPAAATL
RWNPTPEQITLLELYRSGTRTPTEQIQIASKLRKYGRIEGKNVFWFQNHKARERLK
RRRREGGALIKPHKDVKDSSSGGHRVDQTKLCPSPHTNRPQPQHELDPASYNKDNANN
EDHGTTEESDQRASEVGKYATWRNLVTSITQQPEEINIDENVNGEEETRDNRTLNLF
VREYQEKTRGLIEKTKACNYCYEYEFMPLKN*
>G1588 (1..2232)
ATGTACCATCCAAACATGTTTGAGAGCCATCATATGTTTCGATATGACCCCAAAGAGTACC
TCTGATAACGACTTGGGAATCACCGGTAGCCGAGAAGATGACTTTGAGACCAAGTCAGGT
ACCGAAGTCACTACTGAGAATCCTTCTGGTGAAGAGCTTCAAGATCCTAGCCAACGTCCC
AACAAAAAGAAGCGTTACCATCGCCACACGCAACGCCAAATTCAAGAGCTCGAATCATTC
TTTAAGGAATGTCCTCATCCAGATGATAAGCAACGAAAAGAGTTGAGCCGTGATCTCAAT
TTAGAGCCTCTTCAAGTTAAGTTTTGGTTCCAAAACAAACGCACACAGATGAAGGCACAA
AGTGAGAGGCATGAGAACCAGATTCTAAAGTCAGACAAATGACAAGCTCAGAGCAGAGAAC
AATAGATACAAAGAAGCTTAAGCAATGCTACATGCCCTAACTGTGGCGGTCCAGCTGCT
ATTGGAGAAATGCTTTTTGACGAACAACATCTCAGGATCGAAAATGCTCGGCTCCGCGAA
GAGATTGATAGGATCTCTGCTATTGCTGCGAAATACGTTGGGAAGCCGTTAGGATCGTCT
TTCGCTCCACTAGCGATCCACGCGCCTTCTCGTTTCGCTTGATCTTGAAGTTGGAACTTT
GGGAACCAAGACAGGCTTTGTAGGAGAAATGTATGGAACAGGGGACATTTTGAGGTCAGTT
TCGATTCCTTCTGAGACTGATAAGCCTATAATCGTGGAGCTAGCGGTTGCAGCTATGGAG
GAACTCGTGAGAATGGCTCAAACCTGGAGATCCTTTATGGCTTTCAACCGATAATTCAGTC
GAGATTCTCAACGAAGAAGAGTATTTTCAGAACGTTTCCGAGAGGAATTGGACCAAAGCCA
TTAGGATTAAGATCAGAGGCGTCAAGACAATCTGCAGTTGTTATAATGAATCACATCAAT
CTCGTTGAGATTCTCATGGATGTGAATCAATGGTCTTGTGTTTTCTCTGGGATTGTGTCA
AGAGCCTTGACACTTGAAGTTCTTTCAACTGGAGTTGCTGGGAACTACAACGGTGCTTTA
CAAGTGATGACAGCTGAGTTTCAAGTTCCATCACCCCTAGTCCCAACGCGTGAGAATAC
TTTGTGAGATACTGCAAACAACACAGTGACGGCTCTTGGGCTGTGGTTGATGTCTCTTTG
GACAGCCTTAGACCAAGTACTCAATCTTAAGAACTAGAAGAAGGCCTTCAGGTTGTCTG
ATTCAGAATTTGCCATATGGTTATTCTAAGGTTACATGGATAGAGCATATGGAGGTAGAT
GATAGATCAGTTCAACATGTATAAACCGTTGGTTTCAGTCCGGTTTAGCTTTTCGGTGCG
AAACGTTGGGTGGCTACACTCGAACGACAATGCGAGCGGCTTGCTAGCTCCATGGCCAGC
AACATTCTCGTGGTGATCTTTCCGTGATAACGAGTCTTGAAGGAAGGAAGATATGTTGAAG
CTAGCTGAGAGAATGGTTATGAGTTTCTGCAGTGGTGTGGCGCGTGCAGTGCACACGCT
TGGACAACAATGTGACAACAGGATCCGATGATGTTTCGGGTATGACCCGCAAGAGTATG
GATGATCCAGGAAGACCTCCGGGTATTGTTCTTAGTGACGCTACTTCATTCTGGATCCCA
GTTGCTCCCAAACGTGTTTTTGATTTCTCCGTGACGAAAATTCAAGAAAAGAGTGGGAT
ATTCTGTCAAATGGAGGTATGGTTTCAGGAAATGGCTCATATAGCCAATGGTCATGAACCT
GGAACTGTGTCTCCTTGCTCCGAGTCAATAGTGGAAGTTCGAGCCAGAGCAACATGTTG
ATTCACAAAGAGAGCTGTACAGATGCATCAGGATCGTATGTGATTTACGCGCCAGTGGAT
ATAGTGGCGATGAATGTGGTTCTAAGCGGTGGAGATCCTGATTACGTGGCGTTGTTGCCG
TCTGGTTTTGCTATTTTACCGATGGTTTCGGTTGGAGGAGAGATGGGAATCAGCATCAG
GAAATGGTTTCTACTACTTCTTCTGGGAGTTGTGGTGGTTTCGCTTTAACCGTTGCGTTT
CAGATTCTGTTGACTCTGTTCTACAGCTAACTCTCACTTGGCTCGGTGGCTACGGTT
AATAGTCTGATCAAATGTACGGTGGAGAGGATTAAAGCTGCTGTTTCTTGTGATGTTGGA
GGAGGAGCGTAG
>G1588 Amino Acid Sequence (domain in AA coordinates: 66-124)
MYHPNMFESHMFDMPKSTSDNDLGITGSREDDFETKSGTEVTENPSGEELQDPSQRP
NKKKRYHRHTQRQIQELESFFKECPHPDDKQKELSRDLNLEPLQVFWFQNKRTQMKAQ
SERHENQILKSDNDKLAENNRYKEALSNATCPNCGGPAAIGEMSFDQHLRIENARLRE
EIDRISAIKAYVVGKPLGSSFAPLAIHAPSRSLDLEVGNFNGQTGFVGMGYTGDILRSV
SIPSETDKPIIVELAVAAMEELVRMAQTGDPWLWSTDNSVEILNEEYFRTFFRGIGPKP
LGLRSEASRQSAVIMNHINLVEILMDVNQWSCVFSGIVSRALTLEVLSTGVAGNYNGAL
QVMTAEFQVPSPLVPTRENYFVRYCKQHSQSWAVVDVSLSLRPSTPILRTRRRPSGCL
IQELPNGYSKVTWIEHMEVDDRSVHNMYKPLVQSGLAFGAKRWVATLERQ CERLASSMAS
NIPGDL SVITSP EGRKSM LKLAERMVMSFCSGVGASTAHAWTTMSTTGSDDV RVMT R KSM
DDPGRPPGIVLSAATSFWIPVAPKRVDFLRDENS RKEWDILSNGGMVQEMAH IANGHEP

GNCVSLLRVNSGNSQSNMLILQESCTDASGSYVIYAPVDIVAMNVVLSGGDPDYVALLP
SGFAILPDGSVGGGDNQHQEMVSTSSGSCGGSLLTVAFQILVDSVPTAKLSLGSVATV
NSLIKCTVERIKAAVSCDVGGGA*

>G1589 (179..2221)

ACCAAACCTCACATAGCAATCACACACATCTCCACAAACACAGCTTGAGATGATCATGAAA
CACGTGCATCCTCAGATCTCTATCAATCCAGCTTGGTGAAAGAGGTCAAGAAATGAAAG
AGAATCAAAGAAAACGACGTCGTTTCATTCTGTGTAAACACTACTAATTATACATAGAT
GGCTGCTTACTTTACGGAAACCCACCGGAGATCTCTGCCGGATCCGACGGTGGTCTTCA
AACGTTGATCCTCATGAATCCAACACTACTTACGTTTCAGTACACCAACAAGACAACGACTC
GAACAACAACAACACAGCAACAATAGCAACAACAACAACACAAAACACAAACACAAACAA
CAACAACAGTAGTTTCGTTTCTCGATTCCCACGCGCGCGCAGCCAAACGCGAGCCAGCA
GTTCTGTCGGAATACCACTCTCAGGTCACGAAGCTGCTTCCATTACAGCCGCCGACAACAT
CTCCGTACTTTCACGTTATCCTCCGCGCGTGCAGTACAGTCTCTACGGTAGCCACCAAGT
GGATCCCACTCACCAGCAAGCCGCGTGTGAGACGCCACGCGCGCAGCAAGGCCCTCTCTTT
AACCCTCTCGTCTCAACAGCAGCAGCAACAGCAACATCATCAACAACACCAGCCTATTCA
CGTCCGATTCCGGATCAGGATCAGCGAAGTATACGCGGTCGGGTCTGGCTCTACAGGATC
GGGGGTAACAAACGGTATAGCTAATCTTGTTAGCTCCAAGTACTTGAAGGCAGCACAAAGA
GCTTCTTGACGAAGTAGTCAACGCTGATTCCGATGACATGAACGCTAAATCCCAACTATT
CTCATCGAAAAAGGGTAGTTGCGGAAATGATAAACCTGTGCGGAGATCATCGGCCGCGC
TGGAGGAGAAGGTTCCGGTGGCGGAGCAGAAGCAGCCGGGAAACGTCCGGTGGAGCTAGG
CACGGCAGAGAGACAAGAAATACAGATGAAGAAAGCAAACTTAGTAACATGCTTCATGA
GGTGGAGCAGAGATATAGACAGTACCACCAGCAGATGCAGATGGTGATCTCTTCGTTCTGA
GCAAGCGGCAGGGATAGGATCAGCGAAGTCATACACGTCGCTAGCATTGAAAACCATATC
AAGACAGTTCCGTTGCTTGAAAGAGGCGATCGCTGGTCAGATAAAAGCGGCCAACAAAGAG
TCTTGGGGAGGAAGATTCAAGTGTCTGGTGTGGGAGGTTTGAGGGGTCGAGGCTCAAGTT
CGTGGACCACCATTGAGACAGCAAAAGAGCTCTTCAACAACCTGGGAATGATTCAACATCC
TTCCAATAATGCTTGGAGACCTCAACGTTGGTCTCCAGAACGAGCCGTCTCAGTTCTCCG
TGCTTGGCTCTTCGAACACTTTCTTCATCCATACCCTAAGGATTTCGGACAAGCACATGCT
AGCTAAGCAACAGGACTCACTCGTAGCCAGGTGTGCAACTGGTTTATAAACGCGAGAGT
TCGGTTATGGAACCAATGGTGGAGAGATGTACATGGAGGAAATGAAGGAGCAGGCAAA
GAACATGGGATCCATGGAAGAGACTCCTTTGGATCAAAGCAACGAAGATTCTGCTTCAAA
GTCAACAAGTAACCAAGAAAAGAGCCCAATGGCGGACACTAATTACCATATGAATCCCAA
TCACAACGGTGACCTAGAAGGCGTCACTGGAATGCAAGGATGCCCAAGAGACTAAGAAC
CAGCGACGAGACAATGATGCAGCCAATAAATGCGGATTTCAGCTCCAACGAGAAGCTCAC
GATGAAGATTCTAGAAGAACGGCAAGGGATAAGATCAGATGGTGGCTACCCCTTTCATGGG
TAATTCGGGCAATACCAATGGATGAGATGTCAAGATTTGATGTAGTCTCAGACCAGGA
GCTCATGGCGCAAGGTACTCAGGAAACAACAATGGCGTGTCCCTCAGCTTAGGTTTACC
TCATTGTGATAGCTTGTCTGTCACGGACCATCAGGGTTTCATGCAGACCCACCATGGGAT
TCCTATAGGGAGAAGAGTGAAAATAGGAGAAACAGAGGAATATGGACCCGCCACCATCAA
TGGTGGTAGCTCGACCACAACCGCACATTTCATCAGCGGCAGCTGCCGCGGCTTACAATGG
GATGAACATACAGAACCAGAAGAGATATGTGGCTCAGTTATTGCCCGACTTCGTTGCATA
AACCCTCTCTCTAGAAGGAGAAACCGAAACAGGTTATTATATACGTTTCTAGTTTTTAA
TTAGTATATAGTTTCTCATACCAATTGAACCAAAACAAGAAACAAAATTTAATTTTAGTCT
TTGGTTATATATGCGCCGACGGGCTACGTCAGGCCCCCTGACGTAGC

>G1589 Amino Acid Sequence (conserved domain in AA coordinates:384-448)

MAAYFHGNPPEISAGSDGGLQTLILMNPTTYVQYTQDNDNSNNNNNSNNNNNTNTNTN
NNNSSVFVLDSDHAPQPNASQQFVGIPLSGHEAASITAADNISVLHGYPYPRVQYSLYGS HQ
VDPHTHQAACETPRAQQGLSLTLSSQQQQQQHHQHQHPHIVGFGSGHGEDIRVSGSGTG
SGVTNGIANLVSSKYLKAAQELLDEVVNADSDDMNAKSLFSSKKGSCGNDKPVGESSAG
AGGEGSGGGAEEAGKRPVELGTAERQEIQMKKAKLSNMLHEVEQRYRQYHQMQMVISSE
EQAAGIGSAKSYTSLALKTISRQFRCLKEAIAAQIKAANKSLGEEDSVSGVGRFEGSRLK
FVDHHLRQORALQQLGMIQHPSNNAWRPQRGLPERAVSVLRAWLFEHFLHPYPKDSKHM
LAKTGLTRSQVSNWFINARVRLWKPMVEEMYMEEMKEQAKNMGSMKTPLDQSNEDSAS
KSTSNQEKSPMADTNYHMNPNHNGDLEGVTGMQCPKRLRTSDETMMPINADFSSNEKL
TMKILEERQGIIRDGGYPFMGNFQYQMDMSRFDVVSQDELMAQRYSGNNNGVSLTLGL
PHCDLSSTDHGFQTHHGIPIGRRVKIGETEEYGPATINGGSSTTTAHSAAAAAAYN
GMNIQNQKRYVAQLLPDFA*

>G160 (38..784)

TCAAATTTGTCATTTGTTTATTCAAATTTTGGAGAAATGGTGAGAAGTACCAAAGGTCG
TCAGAAAATAGAGATGAAAAAATGGAAAACGAAAGCAACCTTCAGGTTACTTTCTCAA
AAGAAGATTCGGTCTTTTCAAAAAAGCTAGTGAACTTTGCACATTAAGTGGTGCAGAGAT
TCTGTTGATTGTGTTCTCTCCTGGTGGGAAAGTGTCTTTTGGCCATCCAAGTGTTC
AGAACTCATTATCGCTTTTTCGAATCCTAACCATAATTCTGCCATTGTCCATCATCAGAA
CAACAATCTCCAATTTGTTGAAACCCGTCGGGATAGAAATATCCAATATCTCAACAATAT
ACTCACTGAGGTGCTGGCAAACAGGAAAAGGAGAAACAGAAGAGAATGGTTTTGACCT
ATTGAAAGAATCCAGAGAAACAAGTAGGAACTGGTATGAAAAAGATGTGAAAGATCTCGA
CATGAATGAAACCAACCAGCTGATATCTGCTCTTCAAGATGTGAAAAAGAACTGGTAAG
AGAAATGTCTCAATATTCTCAAGTAAATGTTTCGAGAATTACTTTGGTCAAAGTTCTGG
CGTGATTGGTGGTGGTAAATGTTGGCATTGATCTTTTGGATCAAAGAAGAAATGCATTCAA
CTATAATCCAAACATGGTGTTCCTCAATCATAACCCACCAATGTTTGGATACAACAATGA
TGGAGTTCTCGTTCGGATATCCAACATGAACACATGTCAAGTTACAACCTCAACCAGAG
CTAGAGTCTGAAGCTAGAAGAACATCCTAATCAATATTGCGTTATTTGGCTATGGTTA
CTGTTAGGATTGTTCTTGTATTGTGAGACTTAAGTTTGTCTTTCTTTAATTTGTTTCA
GTTGGTTGGTTTTTCATTTTATTTCGTCGTTTGTCTTTGTTTTGTTTGGATATTTTGTGTA
TCCAGAATAAATTTATTTATCCTTTAAAAA

>G160 Amino Acid Sequence (domain in AA coordinates: 7-62)

MVRSTKGRQKIEKMKMENESNLQVTFKRRFGLFKKASELCTLSGAEILLIVFSPGGKVF
SFGHPSVQELIHRFSNPNNHSAIVHHQNNNLQLVETRPDRNIQYLNILTEVLANQBEKEK
QKRMVLDLLKESREQVGNWYEKDVKDLDMNETNQLISALQDVKKLVREMSQYSQVNVVSQ
NYFGQSSGVIGGGNVGIDLFDQRRNAFNYPNPMVFPNHTPPMFGYNNNDGVLVPISNMNYM
SSYNFNQS*

>G1636 (19..666)

GAGTAATCATCAACGATTATGGCGTCAAGTCAGTGGACGAGGTCGGAGGATAAGATGTTT
GAGCAAGCTTTGGTCTTTTTCCTGAAGGATCTCCTAATCGGTGGGAGAGAATCGCTGAT
CAGCTTCATAAATCTGCTGGTGAAGTTAGGGAGCATTACGAGGTCCTTGGTTCATGATGTT
TTCGAGATTGATTCTGGTTCGAGTTGATGTCCCTGATTACATGGATGACTCGGCGGGCTGCC
GCGCGGGGTTGGGATTCCGCTGGTCAGATCTCTTTGGGTCTAAACATGGCGAGAGTGAA
CGCAAAGAGGAACTCCTTGGACAGAGAACGAACACAAATGTTTCTGATCGGATTAAAG
AGATATGGTAAGGGAGATTGGAGGAGTATCTCGAGAAACGTTGTGGTGACGAGGACACCG
ACGCAAGTCGCGAGTCACGCTCAGAAGTATTTCTGAGACAGAACTCGGTGAAGAAGGAG
AGGAAAAGGTCGAGCATCCATGATATAACTACGGTTGATGCTACTTTGGCTATGCCTGGG
TCTAACATGGACTGGACTGGCCAAACCGGGAGTCCTGTTACGGCGCCGAGCAGCAACAG
ATTATGTCTGAGTTCCGTCAGCAATTGAATCCTGGTCATTTCGAGGATTTTGGGTTTCGG
ATGTGATG

>G1636 Amino Acid Sequence (domain in AA coordinates: 100-165)

MASSQWTRSEDKMFQALVLFPEGSPNRWERIADQLHKSAGEVREHYEVLVHDFEIDSG
RVDVPDYMDDSAAGWDSAGQISFGSKHGESERKRGTPTWENEHKLFLIGLKRYGKGD
WRSISRNVVVTRPTQVASHAQKYFLRQNSVKKERKRSSIHDITTVDATLAMPGSNMDWT
GQHGSVPVQAPQQQIMSEFGQQLNPGHFEDFGFRM*

>G1642 (1..1077)

ATGGGTCATCACTCATGCTGCAACAAGCAAAAGGTGAAGAGAGGGCTTTGGTCACCTGAA
GAAGACGAAAAGCTCATCAACTACATCAATTCATATGGCCATGGATGTTGGAGCTCTGTT
CCTAAACATGCAGGTTTGCAGAGATGTGGAAGAGTTGTAGATTAAAGATGGATAAATTAT
CTAAGACCTGATCTTAAACGTGGAAGCTTCTCTCCTCAAGAAGCTGCTCTTATCATTGAG
CTTCACAGCATTCTTGGTAACAGATGGGCTCAAATTGCTAAACATCTACCTGGAAGAACA
GATAACGAGGTCAAGAATTTCTGGAACCTCGAGCATTAAAAAGAAGCTCATGTCTCACCAT
CATCAGGTCATCATCATCATCTCTCTTCCATGGCGAGTTTGCTCACAAACCTTCCT
TATCAATGGATTCAACCCCTACTACAGTCGACGATGAAAGTTCAAGATTCATGTCCAAT
ATCATCACAAACCTAACCCTAATTTTCATCACTCCAAGCCATCTCTCTCTCTCTCTCTCT
CATGTTATGACCCCATGATGTTCCCAACCTCTAGAGAAGGAGATTTCAAGTTTCTAACC
ACAAACAACCCAAACCAATCTCATCACCATGATAATAACCATTAACAACCTCGACATT
TTGTCACCACACCAACTATAACAATCATCATCAACCTTCACTTTCTTCTTGTCTCAT
GATAATAATCTCCAATGGCCAGCGTTACCAGATTTCCAGCGAGTACCATTCTGGTTTC
CAAGAAACCTTCAAGATTATGATGATGCTAATAAACTCAACGTGTTTGTGACACCATTCT

AACGATAATGCCAAAAAGTTATTATGTGGAGAAGTTCTCGAAGGCAGAACTACTATCTTCC
TCCTCACC AATTTT CACAAGATCACGGCCTTTTCTTCCCACCACGTACAAC TTTCAAATG
ACTTCTACGAGTGATCATCAACATCATCATCGAGTGGACTCATAATCAATCACATGATC
ATACCATCATCATCTCATCGTCGCCAATCTCTTGTGGACAGTACGTCATAACTTAA
>G1642 Amino Acid Sequence (domain in AA coordinates: TBD)
MGHHS CCNKQKVKRGLWSPEEDEKLIN YINSYGHGCWSSVPKHAGLQRCGKSCRLRWINY
LRPDLKRGFSFPQEAALIELHSILGNRWAQIAKHLPGRTDNEVKNFWNSSIKKKLMSSH
HHGHHHHHLLSSMASLLTNLPYHNGFNPTTVDESSRFMSNIITNTNPNFITPSHLSLPSP
HVMTPLMFPTSRGDFKFLTNPNQSHHHDNNHYNLDILSPTPTINNHHQPSLSSCPH
DNNLQWPALPDFPASTISGFQETLQDYDDANKLVFVTPFNDNAKKLLCGEVLEGKVLSS
SSPISQDHGLFLPTTYNFQMTSTSDHQHHRVDSYINHMIIPSSSSSSPISCGQYVIT*

>G1747 (1..777)
ATGAAAATGATGCAAGAGGAGGGAACCGAAAAGGTCCATGGACAGAACAGGAAGACATA
CTTCTGGTAAATTTTGTTCACCTTATTTGGAGATCGACGATGGGATTTTATAGCAAAAAGTA
TCAGGTTTGAACAGAACAGGAAGAGTTG CAGGCTAAGATGGGTTAATTACCTACATCCT
GGTCTCAAACCTGGAAGATGACGCCTCAAGAAGAGCGCTCGTCTCTTGAGCTTCACGCT
AAGTGGGGAACAGGTGGTGCAGAAATAGCCCCGAAAATTGCCGGGACGAAACGGATAACGAG
ATAAAGA ACTACTGGAGGACTCATATGAGGAAGAAAGCTCAAGAAAAGAAGCGTCTCTGTT
TCCCCAACTTCTCTATTTTCCAAC TGCAGCTCGTCATCTGTGACCACTACCACCACCAAT
ACTCAAGATACATCGTGCCACTCGCGTAAATCTTCAGGGGAAGTGAGCTTTTACGACACT
GGAGGTTCCCGATCCACTAGAGAGATGAATCAAGAAAACGAAGACGTGTACTCGTTGGAT
GATATATGGAGAGAGATTGATCACTCAGCAGTAAACATAATAAAACCGGTTAAAGACATC
TACTCAGAACAAAGCCATTGCTTAAAGTTACCCAAATCTAGCTTCACCATCATGGGAAAGC
TCATTGGATTCTATATGGAACATGGATGCAGATAAAAGTAAGATATCGTCTTACTTTGCA
AATGATCAGTTTCTTTCTGTTTCCAACACAGTAGATCACCATGGTCTGTCAGGTTAA

>G1747 Amino Acid Sequence (domain in AA coordinates: 11-114)
MKMMQEEGNRKGPWTEQEDILLVNFVHLFGDRRWDFIAKVSGLNRTGKSCRLRWVNYLHP
GLKRGKMTPOEERLVLELHAKWGNRWSKIARKLPGRTDNEIKNYWRTHMRKKAQEKRPV
SPTSSFSNCSSSSVTTTTNTQDTSCHSRKSSGEVSFYDTGGSRTREMNQENEDVYSLD
DIWREIDHSAVNIIPVKDIYSEQSHCLSYPNLASPSWESSLDSIWNMDADKSKISSYFA
NDQFPFCFQHSRSPWSSG*

>G1749 (59..535)
CAACACTTCTCAGTGACCGTGAGCAACGAATTATTTTCAGTTCAACGACTCCGCGGAAAT
GGAAAATTCAGAAAATGTTCCCTCTTACGATCAAAACATCAATTTCACTCCTAATTTGAC
GAGAGATCAAGAACATGTGATCATGGTCTCTGCTTTGCAACAAGTAATATCCAACGTCGG
AGGTGACACGAACCTCGAATGCATGGGAAGCTGATCTTCACCTTTGAACGCTGGCCCTTG
TCCTCTTTGTAGTGTACCCGGCTGCTACGCTTGCCTCTTCCACGACACGAGGCGGATAAT
TAAGAAGGAGAAGAAGCACAAAGGAGTGAGGAAAAAACCATCAGGTAAATGGGCGGCGGA
GATATGGGATCCGAGTTTGAAGTAAGGAGATGGCTTGAACGTTTCCAACAGCGGAGAT
GGCGGCTAAGGCTTACAACGATGCGGCGGCTGAGTTTGTGCGGAAGAGATCAGCAAGACG
TGGCACAAAGAACGGAGAGGAAGCATCTACCAAGAAGACGACTGAGAAAAATTACGGAG
AAGGAGCACGTATAGAAAGGCAGGAAGAGGCATCTTACTTGCTTCACAAGTAAATCAGAA
TTTTTTTTGAAAAGTAAAAACGTTATTTTGTGTTGTAATAAAATAAAGTAAAAACAAATAT
TGCTAACGCAAGACTTATCAAGTTCAGTCGTGACTGTGAGTGTGTTTTATGTATCTTAC
TTCATTTTTTTGTCTTTCAATTGTGTGTGTGTGTGT

>G1749 Amino Acid Sequence (conserved domain in AA coordinates: 84-155)
MENSENVPSYDQINIFPNLTRDQEHVIMVSALQQVISNVGGDTNSNAWEADLPPLNAGP
CPLCSVTGCGYCVFPRHEAIIKKEKKHKGVRKKPSGKWA AEIWDPSLKVRRWLGTFTAE
MAAKAYNDAAA EFGVRRSARRGTKNGEEASTKKTT EKN*

>G1751 (117..923)
AAACACAAACAAAACATATTTTCAATCTCCAGGTGCTTTACACCAACAGAGTCGCAAG
AAAACAAAACCAAACCTCGATTAGTTTGACAGAGAAGGAATCGAGAGTCGGGTATGC
ATTATCCTAACAACAGAACCGAATTCGTGCGAGCTCCAGCCCCAACCCGGTATCAAAGG
AGCAGTTGTACCCGAGCAAGAGCTTTCAGTTATTGTCTCTGCTTTGCAACACGTGATCT
CAGGGGAAAACGAAACGGCGCGGTGTCAGGGTTTTTCCAGTGACAGCACAGTGATAAGCG
CGGGAATGCCTCGGTTGGATTGACACACTGTCAAGTCTGTAGGATCGAAGGATGTCTCG
GCTGTAAC TACTTTTTCCGCCAAATCAGAGAATTGAAAAGAATCATCAACAAGAAGAAG

AGATTACTAGTAGTAGTAACAGAAGAAGAGAGAGCTCTCCCGTGGCGAAGAAAGCGGAAG
GTGGCGGGAAAAATCAGGAAGAGGAAGAACAAGAAGATGGTTACAGAGGAGTTAGGCAAA
GACCTTGGGGAAAAATTTGCAGCTGAGATCAGAGATCCTAAAAGAGCCACACGTGTTTGGC
TTGGTACTTTTCGAAACCGCCGAAGATGCGGCTCGAGCTTATGATCGAGCCGCGATTGGAT
TCCGTGGGCCAAGGGCTAAACTCAACTTCCCCTTTGTGGATTACACGTCTTCAGTTTCAT
CTCTGTTGCTGCTGATGATATAGGAGCAAAGGCAAGTGAAGCGCCAGTGTGAGCGCCA
CAGATTGAGTTGAAGCAGAGCAATGGAACGGAGGAGGAGGGGATTGCAATATGGAGGAGT
GGATGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
CAATGCTGATATGTTCCAGTGATAAATGAGCTCTTTCTTGTGGCGTTTTTTGGAGTTA
AGTGCAAGAAGAGATTGACACTGTGGCTTGTTTAAAGTGAACAAGAACAAGAAAGCATGT
AATTAGTAGTCTCATTCTTTTGTGTTGTGGTCAATTCTATGTTTATCTCATATAAAATCTG
AGTTAAACCTATCTGAGGAGAGAGTAAATAAAGAGGTTAAGAA

>G1751 Amino Acid Sequence (domain in AA coordinates: TBD)
MHYPNNRTEFVGAPAPTRYQKEQLSPEQELSVIVSALQHVISGENETAPCQGFSSDSTVI
SAGMPRLSDSTCQVCRIEGCLGCNYFFAPNQRIEKNHQEETSSSNRRRESSPVAKKA
EGGGKIRKRKNKKNKYRQVRQRPWGKFAAEIRDPKRATRVWLGTFTAEADAARAYDRAAI
GFRGPRAKLNFPFVDYTSVSSPVAADDIGAKASASVSATDSVEAEQWNGGGGDCNME
EWMNMMMMMDFGNGDSSDSGNTIADMFAQ*

>G1752 (25..756)
AAAAAAAAAAAAAAAAAACTTATGGAATATTTCCCAATCTTCCATGTATTCTCTCCA
AGTTCTTGAGCTCATCACAAGAATCACTCTTATGGAACGAGAGCTGTTTCTTGATCAA
TCATCTGAACCTCAAGCCTTCTTTTGCCTAATTATGATTACTCCGATGACTTTTTCTCA
TTTGAGTCACCGGAGATGATGATTAAGGAAGAAATTCAAAACGGCGACGTTTCTAACTCC
GAAGAAGAAGAAAAGGTTGGAATTGATGAAGAAAGATCATACAGAGGAGTGAGGAAAAGG
CCGTGGGGGAAATTTGCAGCGGAGATAAGAGATTCAACGAGGAATGGAATTAGGGTTTGG
CTCGGGACATTTGACAAAAGCCGAGGAAGCCGCTCTTGCTTATGATCAAGCGGCTTTCGCC
ACAAAAGGATCTCTTGCAACACTTAATTTCCCGGTGGAAGTGGTTAGAGAGTCGCTAAAG
AAAATGGAGAATGTGAATCTTCATGATGGAGGATCTCCGTTATGGCCTTGAAGAGAAAA
CATTCTCTTCGAAACCGGCCTAGAGGGAAAAAGCGATCCTCTTCTTCTTCTTCTTCT
TCTAATCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
AAGCAGAGTGTGTGAAGCAAGAAAGTGGTACACTTGTGGTTTTTGAAGATTTAGGTGCT
GAGTATTTAGAACAACCTTCTTATGAGCTCATGTTGATCTTGTAAATTGATTTCAGCAAAAG
CCACTATTAAACTTTAATTTTGTGATAATTAATCTTGAATTTGTTTTGTTTCTTCTGCA
ATTTCTTTGGTTCTCTTATTTTGTGTTGTGATCCAAATGAAATTATTGAAGAGATG
GTGATGTTAAAGTGATATATATATAAAAAAAAAA

>G1752 Amino Acid Sequence (domain in AA coordinates: TBD)
MEYSQSSMYSSPSSWSSSQESLLWNESCFLDQSSEPQAFPCPNYDYSDDFFSFESPEMMI
KEEIQNGDVSNSEEEKVGIDEERSYRGVRKRPWGKFAABIRDSTRNGIRVWLGTDFDKAE
EALAYDQAAFATKGSLATLNFPEVVRESLKKMENVNLHDGSPVMALKRKHSLRNRPR
GKKRSSSSSSSSSSSSSSSSSSSTSRSSSKQSVVKQESGTLVVFEDLGAEYLEQLLM
SSC*

>G1763 (33..977)
GTACATTTTTTTTTTGTATTTTCAGGAACTCCGATGGCGGATCTCTTCGGTGGTGGCCACG
GCGCGGAGCTTATGGAAGCACTTCAACCTTTTTTACAAAAGTGCTTCCACGTCTGCTTCAA
ATCCTGCGTTTTCGCTCCTCAAACGATGCGTTTTCGCTCTGCCCCAAACGACCCATTTCTT
CTTCTTCTTACTATAATCCTCATGCATCTTTCTTCCCTTACATTCACAAACCACTTACC
CGGATATTTATTCTGGATCCATGACCTATCCATCTTCATTGCGGTTCGGATCTTCAACAAC
CCGAAAACCTACCAATCTCAGTTCCATTACCAAAACACTATCACTTACACTACCAAGACA
ACAACACTTGCATGCTCAACTTCATTGAGCCGAGCCAACCGGATTTATGACCCAACCGG
GTCGAGTTTCGGGTTTCGGTTTCAAAACCGGCTAAGCTCTATAGAGGAGTGAGGCAAAGAC
ATTGGGGAAAATGGGTTCGCGGAGATCCGTTTACCCAGGAACCGAACCCGACTTTGGCTCG
GAACATTCGACACGGCTGAAGAAGCCGCTTGGCTTATGATCGCGCCGCTTTAAGCTTC
GTGGTGACTCGGCTCGGCTTAACCTTCCAGCTCTCCGATACCAAAACCGGCTCGTCTCCGT
CTGACGTTGGCGAATACGGACCTATTCAAGCTGCCGTTGACGCCAAGCTAGAAGCCATAT
TAGCTGAGCCGAAGAATCAGCCGGGCAAAACGGAGAGGACGTCGAGGAAACGAGCTAAAG
CCGCGGCTTCTTCACTGAGCAGCCGTCAGCGCCACAACAACATTCCGGGTTCGGGTGAAA
GTGATGGGTTCGGGTTACCGACTTCGGATGTTATGGTGAGGAGATGTGCCAAGAGCCAG

GCGGCAAACGGAGCAATCTTGAAGCAGTAGACGGAGAGGCAAAGATCCACATCGTTGAC
ATAAGCTCCACGTTTTCACCTCAATGGCCGACTCTTCTAGAAGCTTTAGCCACAAGATCA
GACGACACGCCCTCACCTAAGGCTAACCACAGTTGTCGTGGCCAACAAGTTTGTCAACGAT
CAAACGGCGTCGCATCGGATGATGAAAGAGATCGGAAACCGAATGGAGAAATTCGCTAGG
CTTATGGGAGTTCTCTTCAAATTAACATTATTCATCACGTTGGAGATTTATCTGAGTTT
GATCTCAACGAACTCGACGTTAAACCAGACGAAGTCTTGGCCATTAACTGCGTAGGCGCG
ATGCATGGGATCGCTTCACGTGAAGCCCTAGAGACGCTGTGATATCGAGTTTCCGACGG
TTAAGACCGAGGATTGTGACGGTCGTAGAAGAAGAAGCTGATCTTGTGCGAGAAGAAGAA
GGTGGCTTTGATGATGAGTTCTTGAGAGGGTTTGGAGAATGTTTACGATGGTTTAGGGTT
TGCTTCGAGTCATGGGAAGAGAGTTTCCAAGGACGAGCAACGAGAGGTTGATGCTAGAG
CGTGCAGCGGGACGTGCGATCGTTGATCTTGTGGCTTGTGAGCCGTCGGATTCCACGGAG
AGGCGAGAGACAGCGAGGAAGTGGTCCGAGGAGGATGAGGAATAGTGGGTTTGGAGCGGTG
GGGTATAGTGATGAGGTGGCGGATGATGTCAGAGCTTTGTTGAGGAGATATAAAGAAGGT
GTTTGGTCGATGGTACAGTGTCTGATGCCGCCGAATATTCCTTTGTTGGAGAGATCAG
CCGGTGGTTTGGGCTAGTGCGTGGCGGCCAACGTAA

>G1767 Amino Acid Sequence (domain in AA coordinates: 255-272)

MDTLFRLVLSLQQQQSDSIITNQSSLSRTSTTTTGGSPQTAYHYNFPQNDVVEECFNFFMD
EEDLSSSSSHHHNNHNNPNTYSPFTTPTQYHPATSSPTSSSTAAAAALASPYSSSGHHN
DPSAFSIPQTPPSFDFSANAKWADSVLLEAARAFSDKDTARAQQILWTLNELSSPYGDTE
QKLASYFLQALFNRMGTSGGERCYRTMVTAATEKTCSESTRKTVLKFQEVSPWATFGHV
AANGAILEAVDGEAKIHVIDISSTFCTQWPTLLEALATRSDDTPhRLRTTVVANKFVND
QTASHRMMEKIGNRMEKFARLMGVPFKNIIHHVGDLSFDLNELDVKPDEVLAINCVGA
MHGIASRGSPRDAVISSFRRLRPRIIVTVEEEDLVGEEEGGFDEFLRGFGCELRWFRV
CFESWEESFPRTSNERLMLERAAGRAIVDLVACEPSDSTERRETARKWSRRMRNSGFGAV
GYSDEVADDVRALLRRYKEGVWSMVQCPDAAGIFLCWRDQPVVWASAWRPT*

>G1778 (1..627)

ATGATGGGATACCAAACAACTCTAATTTCTCCATGTTTTTTTCTCGGAAATGACGAC
CAAAACCACCACAACCTACGATCCTTATAATAATTCTCTTCATCAACTTCTGTTGATTGC
ACTCTCTCACTTGAACACCCTCTACTCGTCTCGACGACCACCATAGATTTTCTCTGCT
AATTCTAACACATCTCCGGCGACTTTTATATTACGGAGGAAACGCTAAGACTTCTTCG
TACAAGAAGGGTGGTGTGCTCATAGCCTACCTCGCCGTTGTGCTAGCTGCGACACCACT
TCAACTCCTCTATGGAGAAACGGACAAAAGGACCTAAGTCGTTATGTAACGCGTGTGGA
ATCCGATTCAAGAAAGAGGAGAGGCGTGCAGCGCCAGAACTTAACGATCTCCGGTGGA
GGTTCATCAGCGGCAGAGTCCAGTAGAGAATTCGTACAACGGAGGTGGAACTATTAC
AGTCATCATCATCACTATGCCCTCGTCTGCGCGTCTGTTGGCTCATCAGAACACACAA
AGAGTTCCATATTTCTCACCGGTTCCGGAGATGGAATATCCCTACGTGGATAACGTCACG
GCTTCTCTTTTATGTCTTGAATTGA

>G1778 Amino Acid Sequence (domain in AA coordinates: 94-119)

MMGYQTNSNFSMFSSSENDQNHNYDPYNNFSSSTSVDCITLSLGPSTRLDHHRFSSA
NSNNISGDFYIHGNAKTSYKKGVAHSLPRRCASCDTTSTPLWRNGPKGPKSLCNACG
IRFKKEERRATARNLTISGGGSSAAEVPVENSYNGGNYSHHHHHYASSSPSWAHQNTQ
RVYPYFSPVPEMEYPYVDNVTASSFMSWN*

>G1789 (108..413)

CAAGGACTCTGCGACATCTGTGCAACATATCATTTCTCAGAATCTCTTTCTTTCTAGG
TTTATTACTACACAAAACCAACATCATCAACTTTAGTTACTAAACAATGGCATCAGGCT
CAATGTCTTTCTTATGGCTCTGGCTCATGGACTGTTAAGCAGAACAAAGCCTTTGAGCGTG
CTCTAGCAGTCTATGACCAAGACACTCCGGACCGTTGGCACAATGTTGCTAGAGCTGTTG
GTGGTAAACACCAAGAGAGCTAAGAGACAGTATGACCTTCTAGTTCGTGACATCGAAA
GCATCGAGAATGGTCACGTGCCATTCCCTGACTACAAGACTACTACAGGAAACAGCAACA
GAGGCAGGCTCGGTGATGAGGAAAAGAGGATGAGAAGCATGAAGTGCAGTGAGACAAGA
AGCAACAAAACCTAAGTACGTATGATCGTCAAAAATAAAGAGAATCACTTCAGAGAGATG
TGTTTTTTTCAATGTCTGACGAATCAATGTTTTTTTCTTGCAATTTCTCATGTTTTTCCC
TAAGAAATGGTTTTTTTTTCGAGGCAACAAAAA

>G1789 Amino Acid Sequence (domain in AA coordinates: 1-50)

MASGSMSSYSGSWTVKQNKAFERLAVYDQDTPDRWHNVARAVGGKTPEEAKRQYDLLV
RDIESIENGHVFPDYKTTTGNNSNRGLRDEEKRMRSMKLQ*

>G1790 (63..1346)

GAAAAAGACTTCACCTTTTTTTTTTACTAATTAATTAGTTTTTTTTTCTCCTTTCCAAAA
CAATGGAGAATTTTCGTCGACGAGAATGGTTTTGCTTCTCTAAACCAAAACATCTTCACAC
GTGATCAAGAACACATGAAAGAAGAGATTTTCCATTGGAAGTCGTCGACCAATCAAAAC
CTACAAGCTTTCTTCAAGATTTTACCATCTTGATCATGATCATCAGTTTGATCATCATC
ATCATCATGGCTCCTCATCTTCACATCCTTTGCTCAGCGTCCAAACTACGCTCTTCTGTAT
TCAATAATGCTCCTTTTCGAGCATTGCTCTTACCAAGAAAACATGGTCGATTTCTATGAAA
CTAAACCAAAATTTGATGAATCATCATCATTTCCAAGCAGTGGAAAACTCATACTTCACTC
GTAATCATCATCATCATCAAGAGATCAATTTGGTCGATGAACATGATGATCCTATGGACT
TGGAGCAAAACAACATGATGATGATGAGGATGATCCCTTTTGATTACCTCCTACAGAGA
CTTTCAAACCTATGAACCTTCGTAATGCCAGATGAAATTTTCATGTGTTTCTGCAGATAATG
ATTGTTATAGAGCAACGAGTTTCAACAAGACCAAAACATTTCTTACACGAAAGTTGTCTT
CTTCTTCTTCATCATCATCATGGAAGAAACCAAAAAGTCAACCTTAGTCAAAGGACAAT
GGACTGCTGAAGAAGACAGGGTACTGATTCAACTCGTGGAGAAGTATGGATTGCGTAAAT
GGTCGCATATCGCTCAAGTGTTACCGGAAGAATCGGGAACCAATGTAGAGAGAGGTGGC
ATAACCATTTGAGACCTGACATTAAGAAAGAAACATGGAGTGAAGAAGAGGACAGAGTGT
TGATAGAATTTACAAAGAGATTGGAAACAAATGGGCAGAGATTGCGAAAAGACTCCCGG
GAAGAACAGAGAACTCGATCAAGAACCATTGGAACGCAACAAAAAGAAGACAATTTCTTA
AAAGAAAATGTAGATCTAAGTATCCAAGACCTTCTCTGTTGCAGGATTACATCAAGAGCT
TGAATATGGGAGCTTTGATGGCTTCTTCTGTTCTGCAAGAGGTAGACGCAGAGAGAGTA
ATAACAAGAAGAAGGATGTTGTTGTTGCGGTTGAGGAGAAGAAGAAGGAAGAGGAGGTGT
ATGGACAAGACAGGATGTGCTGCTGAATGTGTGTTTACTGATGATTTTGGATTCAATGAGA
AGCTGCTTGAGGAAGGATGTAGCATTGACTCTTTGCTTGATGACATTCCTCAGCCTGACA
TTGATGCTTTTGTTCATGGGCTCTGATTTGTATTTTATTCTGCTTGTTCAGTTTGT
TGTTTTTGTATTGTCTTTTATACGAGACAGATTCCACCAAACTTCAATAATTTGAAAAG
ATATAAAATATTTGCTTTTAAAAA

>G1790 Amino Acid Sequence (conserved domain in AA coordinates:217-316)

MENFVDENGASLNQNIFFTRDQEHMKEDFPFEVVDQSKPTSLQDFHHLDDHHDHDFHHH
HHGSSSSHPLLSVQTSSCINNAPFEHCSYQENMVDFYETKPNLMNHHHFQAVENSFYFTR
NHHHHQEINLVDEHDDPMDLEQNNMMMRMIPFDYPPTETFKPMNFVMPDEISCVSADND
CYRATSFNKTKPFTRKLSSSSSSSSWKETKSTLVKGQWTAEDRVLIQLVEKYGLRW
SHIAQVLPGRIGKQCRERWHNHLRPDIKKETWSEEDRVLIEFHKEIGNKWAIEIAKRLPG
RTENSIKNHWNATKRRQFSKRKCRSKYPRPSLLQDYIKSLNMGALMASSVPARGRRRESN
NKKKDVVVAVEEKKKEEVYQDRIVPECVFTDDFGFNEKLLEEGCSIDSLDDIPQDPI
DAFVHGL*

>G1791 (36..455)

ATGTACATGCAAAACAAAACCTTAAAGCTTTCATGGAACGTATAGAGTCTTATAACA
CGAATGAGATGAAATACAGAGGCGTACGAAAGCGTCCATGGGGAATATGCGGCGGAGA
TTCCGCACTCAGCTAGACACGGTGCTCGTGTGTTGGCTTGGGACGTTTAAACACAGCGGAAG
ACGCGGCTCGGGCTTATGATAGAGCAGCTTTCGGCATGAGAGGCCAAAGGGCCATTCTCA
ATTTTCTCAGAGTATCAAATGATGAAGGACGGTCCAAATGGCAGCCACGAGAATGCAG
TGGCTTCCTCGTCTCGGGATATAGAGGAGGAGTGGTGGTGATGATGGGAGGGAAGTTA
TTGAGTTCGAGTATTTGGATGATAGTTTATTGGAGGAGCTTTTAGATTATGGTGAGAGAT
CTAACCAAGACAATTTGAACGACGCAAAACCGCTAGATCATCACTACTTACTTACAGTGTA
ATGTTTTTGGAGTAAAGAGTAATAATCAATATAATATACTTTAGTTTAGGAAAAA
AAAAA

>G1791 Amino Acid Sequence (domain in AA coordinates: TBD)

MERIESYNTNEMKYRGVRKRPWGKYAAEIRDSARHGARVWLGTFTNTAEDAARAYDRAAFG
MRGQRAILNFPHEYQMMKDGPNGSHENAVASSSSGYRGGGGDDGREVIEFEYLDSDLLE
ELLDYGERSNQDNCNDANR*

>G1793 (59..1783)

AGTGATTATTGATTAAACCAACCAAAATAAACAGATTTGACTCAAAAAGAAGAAAAT
GAATTCTAACAACTGGCTTGGCTTTCTCTTACCGAACAACCTTCTTTGCTCTCTCA
TGAATACAACCTTGGCTTGGTCAGCGACCATATGACAACCTTTTCAAACACAAGAGTG
GAATATGATCAATCCACAGGTGGAGGAGGAGATGAAGGAGGAGAGGTTCCAAAAGTGGC
CGATTTTCTCGGTGTGAGCAAACCGGACGAAAACCAATCCAACCACCTAGTAGCTTACAA
CGACTCAGACTACTTCCATACCAATAGCTTGATGCCTAGCGTCCAATCAAACGATGT
CGTTGTAGCAGCTTGTGACTCCAATACTCCTAACACAGTAGCTATCATGAGCTTCAAGA

GAGTGCTCACAATCTACAGTCACTTACTTTGTCCATGGGGACCACCGCTGGTAATAATGT
TG TAGACAAAGCTTCACCATCCGAGACCACCGGGGATAACGCTAGCGGTGGAGCACTAGC
CGTTGTTGAGACGGCCACGCCAAGACGTGCATTGGACACTTTCGGACAACGAACCTCGAT
CTATCGTGGTGTCAAGACATCGATGGACTGGTCGATATGAGGCTCATCTATGGGATAA
TAGTTGTAGAAGGGAAGGCCAGTCTAGGAAAGGAAGACAAGTTTACTTGGGTGGATATGA
CAAAGAAGATAAAGCAGCAAGATCATATGATCTAGCTGCACCTAAGTACTGGGGTCCCTTC
AACTACTACTAATTTCCCATTTACAACTACGAGAAAGAGTAGAGGAAATGAAGCACAT
GACGAGACAAGAGTTCTGTGGCTGCCATTAGAAGGAAAAGTAGTGGATTTTCGAGAGGCGC
TTCGATGTATCGAGGAGTTACAAGGCATCACCAACATGGAAGATGGCAAGCAAGGATCGG
CCGAGTCGCCGAAACAAAGACCTCTACTTGGGAACCTTTAGCACTGAGGAAGAAGCAGC
AGAAGCTTACGATATAGCTGCAATAAAGTTTAGAGGACTTAATGCAGTGACCAACTTCGA
GATCAACCGGTACGACGTGAAAGCCATTCTAGAGAGTAGCACTCTTCCCATCGGAGGAGG
CGCAGCTAAACGGCTCAAAGAAGCTCAAGCTCTTGAGTCTTCAAGGAAACGCGAGGCGGA
GATGATAGCCCTTGGTTCAAGTTTCCAGTACGGTGGTGGCTCGAGCACAGGCTCTGGCTC
CACCTCATCAAGACTTCAGCTTCAACCTTACCTCTAAGCATTCAACAACCATTAGAGCC
TTTTCTATCTCTTCAACAATGACATCTCTCATTACAACAACAATGCTCAGGATTC
CTCTCTTTTAATCACCATAGCTATATCCAGACACAACCTTCATCTCCACCAACAGACCAA
CAATTACTTGCAGCAACAGTCGAGCCAGAACCTCAGCAGCTCTACAATGCGTATCTTCA
TAGCAATCCGGCTCTGCTTCATGGACTTGTCTCTACCTCTATCGTTGACAACAATAATAA
CAATGGAGGCTCTAGTGGGAGCTACAACACTGCAGCATTTCTTGGGAACACGGTATTGG
TATTGGGTCCAGCTCGACTGTTGGATCGACCGAGGAGTTTCCAACCGTTAAACAGATTA
CGATATGCCTTCCAGTGATGGAACCGGAGGGTATAGTGGTTGGACAGTGAGTCTGTTCA
GGGGTCAAACCTGGTGGTGTTTTCACTATGTGGAATGAGTAAACAAGGATCTCTTTCTT
GCGGCACAAGGAATGGGT

>G1793 Amino Acid Sequence (conserved domain in AA coordinates:179-255, 281-349)

MNSNNWLGFPPLSPNSSLPPHEYNLGLVSDHMDNPFQTQEWNMINPHGGGGDEGEVVKV
ADFLGVSKPDENQSNHLVAYNDSYFHTNSLMPVSQSNVVAACDSNTPNNSSYHELQ
ESAHLNQLSLTSMGTTAGNNVVDKASPSSETTGDNASGGALAVVETATPRRALDTFGQRTS
IYRGVTRHRWTGRYEHLWDNSCRREGQSRKGRQVYLGGYDKEDKAARSYDLAALKYWGP
STTTNFPITNYEKEVEEMKHMTRQEFVAAIRRKSSGFSRGASMYRGVTRHHQHGRWQARI
GRVAGNKDLYLGTFTSEEEAEAYDIAAIKFRGLNAVTFEINRYDVKAILLESSTLPIGG
GAAKRLKEAQALESRRKREAEMIALGSSSFQYGGGSSTGSGSTSSRLQLQPYPLSIQPPLE
PFLSLQNDISHYNNNNAHDSSSFNHHSYIQTLHLHQQTNNYLQQSSQNSQQLYNAYL
HSNPALLHGLVSTSIVDNNNNNGSSGSYNTAAFLGNHIGIGSSSTVGSSTEEFPTVKTD
YDMPSSDGTGGYSGWTSSESVQGSNPGGVFTMWNE*

>G1795 (27..422)

ACAAACACGCAAAAAGTCATTAATATATGGATCAAGGAGGTGCGAGGTGTCGGTGCCGAGC
ATGGAAAGTACCGGGGAGTTCGGAGACGACCTTGGGGAAAATATGCAGCAGAGATACGAG
ATTTCGAGGAAGCACGGTGAACGTGTGTGGCTTGGAAACGTTTCGATACGGCAGAGGAAGCGG
CTAGAGCCTATGACCAAGCTGCTTACTCCATGAGAGGCCAAGCAGCAATCCTTAACCTCC
CTCATGAGTATAACATGGGGAGTGGTGTCTCTTCTTCCACCGCCATGGCTGGATCTTCCT
CCGCTCCGCTCCGCTTCTTCTTCTTAGGCAAGTTTGAATTGAGTACTTGGATG
ATAGTGTTTTGGAGGAGCTCCTTGAGGAAGGAGAGAAACCTAACAGGGCAAGAAGAAAT
GAGCGAGATATAATTCATGATTATTTCTAA

>G1795 Amino Acid Sequence (domain in AA coordinates: 12-80)

MDQGGRGVGAEHGKYRGVRRRPWGKYAAEIRDSRKHGERVWLGTFTDABEAAARAYDQAA
SMRQQAAILNFPHEYNMGSGVSSSTAMAGSSSASASASSSRQVFEFEYLDLDSVLEELLE
EGEKPNKGKKK* -

>G1800 (61..894)

CCATTATCATATCCTCTTCTTCTTCTTCACTATCAATCTTCTTCTCCACTACAACACAA
ATGGAGAAATCATCTCAATGAAACAATGGAAGAAGGGTCTGCTCGGGGTAAAGGCGGT
CCACAAAACGCTCTTTGTAGTACCGTGGAGTCAGGCAAGGACTTGGGGCAAATGGGTG
GCTGAGATCAGAGAGCCCAAGAAGAGGGCAAGACTTTGGCTTGGCTTTTCGCTACAGCT
GAAGAAGCAGCTATGGCTTATGATGAGGCTGCCTTGAAACTCTATGGGCACGACGCATAC
CTCAACTTACCTCATCTTTCAGCGGAATACAAGACCTTCTCTGAGTAACTCTCAGAGGTC
AAATGGGTACCTTCAAGGAAGTTTATATCTATGTTTCTTTCATGTGGTATGCTAAACGTG
AATGCTCAGCCTAGTGTTCACATAATCCAGCAAAGACTAGAAGAACTCAAGAAAACCTGGA

CTTTTATCTCAATCCTATTCTTCTAGTTCTTCTCCACCGAATCAAAAACTAATACTAGC
TTTCTTGATGAGAAGACCAGCAAGGGAGAAAACAGACAATATGTTTCAAGGTGGTGATCAG
AAGAAACCAGAGATCGACCTGACCGAGTTTCTTTCAGCAACTAGGAATCTTGAAGGATGAA
AATGAAGCAGAACCAAGTGAGGTAGCAGAGTGTCTATCCCTCCACCATGGAACGAGCAA
GAAGAACTTGAAGTCTTTTCAGAACTGAGAATTTAGCTGGGATACCTGATCGAGATG
CCAAGAAGTGAAACCACAACATATGCAATTTGACTCCAGCAACTTCGGAAGCTATGATTTT
GAGGATGATGTATCCTTCCCTTCCATCTGGGACTACTACGGAAGCTTAGATTGAGTAAAA
GCAATTTAAGGTAGATCAAGATTGAGAAGTACACAAATGGTTTTGGATTTAGTGTAGCGT
TTTGGAAAAGAGACATAGGTAGTGAGAGTGACAGTCTTTTATTATGCAGCAATAAAGTGAG
TCAGTGTAACACCGAGTTGTTTCGCTTTTTTTGGTATATTAATGAAGCATGTTTATTTTT
CGCTAAAAA

>G1800 Amino Acid Sequence (domain in AA coordinates: TBD)
MEKSSSMKQWKKGPARGKGGPQNALCQYRGVVRQRTWGWVAEIREPKKRRLWLGSFATA
EEAAMAYDEAALKLYGHDAYLNLPHLQRNTRPSLSNSQRFKWPSPRKFISMFPSCGMLNV
NAQPSVHI IQQRLEELKKTGLLSQSYSSSSSTESKNTSFLDEKTSKGETDNMFEGGDQ
KKPEIDLTEFLQQLGILKDENEAEPSVABCHSPPPWNEQEETGSPFRTENFSWDTLIEM
PRSETTTMQFDSSNFGSYDFEDDVSPFSIWDYYGSLD*

>G1806 (1..1356)
ATGCAGAGCAGCTTCAAAACCGTTCCCTTCACTCCTGATTTCTACTCTCAATCCTCTTAC
TTCTTCAGAGGAGATAGTTGTCTTGAGGAGTTTCATCAACCAGTCAATGGTTTTCCACCAT
GAAGAAGCTATCGATTTAAGTCCAAATGTCACTATTGCTTCAGCTAACTTACACTACACG
ACGTTTGATACGGTTATGGATTGTGGTGGTGGTGGTGGTGGCTTGAGGGAGAGACTTGAA
GGAGGAGAAGAGGAGTGTTTGGACACAGGGCAATTAGTGTAACAGAAAGGGACAAAGATTA
GTAGGAGGAGGAGTAGGAGAAGTGAACAGCAGTTGGTGTGATTTCGGTTTCAGCTATGGCT
GATAACAGTCAACATACTGACACTTCCACAGATATTGATACTGATGACAAGACTCAGTTG
AATGGAGGTATCAAGGGATGCTATTGGCTACAAATTTGTTTCAGATCAATCCAATGTGAAA
TCTAGTGATCAAAGGACACTTCGTCGACTTGCTCAGAACCGGGAGGCTGCTAGGAAAAGT
CGGTTGAGGAAAAGGCCATATGTTTCAGCAACTTGAGAATAGTCAATCAGGCTTGACACAG
CTAGAGGAAGAGCTCAAAAGAGCTCGCCAACAGGGATCTTTGGTTGAAAGAGGAGTTTCA
GCGGATCACACGCATTTGGCAGCAGGAAATGGTGTCTTTTCATTTGAATTGGAATATACA
CGTTGGAAGGAGGAACATCAAAGAATGATCAACGACTTAAGATCGGGTGTGAATTCGCAG
TTAGGTGACAACGATCTACGCGTTCTAGTGGATGCTGTGATGAGTCACTATGATGAAATA
TTCAGGCTAAAGGGAATTGGCACTAAAGTTGAAGTCTTTCATATGCTCTCAGGCATGTGG
AAGACACCTGCCGAGAGATTTTTTCATGTGGTTAGGTGGATTTAGATCATCAGAGTTACTT
AAGATATTGGGGAACCATGTGGATCCATTGACGGACCAGCAGTTGATAGGCATTTGCAAC
CTTCAGCAATCGTCTCAACAAGCAGAGGATGCATTGTCAAGGCATGGAAGCTCTACAA
CAATCACTTCTCGAGACGCTTTCTTCTGCTTCTATGGGTCCAACTCTTCAGCAAAATGTT
GCAGATTATATGGGTATATGGCTATGGCTATGGGCAAACTTGGCACTCTTGAAAACCTTC
CTTCGCCAGGCTGATTTATGAGGCAACAACTCTGCAACAGCTTCACAGAATTCTCACC
ACACGACAAGCTGCTCGCGCCTTTTTGGTTCATCCACGATTATATTCTCGGCTTAGAGCA
CTTAGCTCTCTATGGTTAGCCAGACCTAGAGACTAA

>G1806 Amino Acid Sequence (domain in AA coordinates 165-225)
MQSSFKTVPFPTDFYSQSSYFFRGDSCLEEFHQPVNGFHHHEAIDLSPNVTIASANLHYT
TFDVTMDCGGGGGGLRERLEGGEEELDTGQLVYQKGTRLVGGGVGEVNSSWCDSVSAMA
DNSQHTDSTDIDTDDKTQLNGGHQGMLLATNCSQSNVKSDDQRTLRLRAQNREARKS
RLRKAYVQQLENSRIQLAQLEELKRARQQGSLVERGVSAADHHLAAGNGVFSFELEYT
RWKEEHQRMINDLRSGVNSQLGDNDRVLVDVMSHYDEIFRLKGIGTKVEVFHMLSGMW
KTPAERFFMWLGGFRSSELLKILGNHVDPLTDQQLIGICNLQSSQQAEDALSQGMALQ
QSLLETLSASMGPNSSANVADYMGHMAMAMGKLTLENFLRQADLLRQOTLQQLHRIILT
TRQAARAFVVIHDYISRLRALSSLWLARPRD*

>G1811 (93..827)
AAAGGAGCATTTGGTATCTCAACAATATTTGCCCTTTCTCTATCTCTCTCATCACTAT
TTGCCATCTCTTTCTCTCTCCCTCTCTTTCAAATGTCAATAAACCAATACTCAAGCGATT
TCCACTACCATTTCTCTCATGTGGCAACAACAGCAGCAACAACAACACCAAACGACG
TCGTGGAAGAAAAAGAAGCTTTTTTCGAGAAACCCCTTAACCCCAAGTGACGTCGAAAAAC
TCAACCGCCTCGTCATCCCAAAACAGCAGCCGAGAGATACTCCCACTAGCGGCCGCCG
CCGCAGACGCCGTGGAGAAAGGACTTCTCCTCTGCTTTGAGGACGAGGAAGGTAAACCAT

GGAGATTGAGATACTCGTACTGGAACAGTAGCCAGAGTTATGTCTTGACCAAAGGCTGGA
GCAGATACGTCAAGGAGAAGCACCTTGACCGCGGAGACGTCGTTCTCTCCATCGACACC
GTTGAGACGGCGGAAGATTCTTCATTGGCTGGAGAAGACGCGGTGACTCTTCTTCTCCT
CCGACTCTTATCGCCATGTTCATCAATCCAATGCCTCGCTCCAATATTATCCTCATGCAGGGG
CTCAAGCGGTGGAGAGCCAAAGAGGCAACTCGAAGACATTAAGACTGTTGCGAGTGAACA
TGGAGTGCCAGCTAGATTGCGACTGGTCCGAGCCATCCACACCTGACGGTCTTAACACAT
ATACAACCAATCACGACCACTTTCATTTCTACCCTCAACAACAACACTATCCTCCTCCGT
ACTACATGGACATAAGTTTCACAGGAGATATGAACCGGACGAGCTAGAAGCCCACAAAGGA
TTAAAAAAAAGCTTCACATCTGGTCTCTGTTATGTTGTCATAGATGTTGATTCCTTAATTT
TACACAAGCTTCATTTTGCAATTATTTAAAGTAAAATCGTATTTTGATTCTCTTTAAATC
TCTCTCAATTTTCACTCTCTCCTTTTCTTCTTATGTATTAGATTCTTTTACATAGCTA
ACACTTGTATAGAGAATTCAAAGTTCTGGCTATTTTCGAAAGTTATCTTTTCTCTTAAAA
AAAAAAA

>G1811 Amino Acid Sequence (domain in AA coordinates: TBD)
MSINQYSSDFHYHSLMWQQQQQQQHQNDDVVEEKEALFEKPLTPSDVGKLNRLVIPKQHA
ERYFPLAAAAADAVEKGLLLCFEDEEGKPPWFRFRYSYWNSSQSIVLTGWSRYVKEKHLDA
GDVVLFRHRSDGGRFFIGWRRRGDSSSSSDSYRHVQSNASLQYYPHAGAQAQVESQRGNS
KTLRLFGVNMECQLDSDWSEPSPTDGSNTYTTNHDQFHFYPQQQHYPPPYMDISFTGDM
NRTS*

>G182 (74..1366)
CGTCGACGATCAGATTCTTGCGTATAGCTGTATATATACACCAAGATACACTCATCATCG
TCATATATAGATTATGTGACGCGTCTCTGAGCTTCTTGACATGGAAAACCTCCAAGGAGA
CTTAACCGACGTCGTACGAGGAATCGGAGGCCACGTGTTATCACCGGAGACTCCTCCCTC
GAACATCTGGCCTCTTCTCTGTACATCCAACACCATCACCGTCAGATCTTAACATAAA
CCCCCTCGGAGATCCGCTTTGTGAGCATGGACGATCCACTCCTCCAAGAACTAAACTCCAT
CACAAACTCCGGCTATTTCTCCACCGTAGGAGATAACAACAACAACATTACACAACAACA
TGGTTTCTTGGTTCCAAAGGTATTTGAGGAGGATCATATAAAGAGTCAATGTAGTATCTT
CCCAAGAATCCGGATCTCGCATAGTAACATCATCCACGATTCTTCTCCGTGTAATCTCC
GGCCATGTGCGGCTACGTTGTGCGAGCCGACGAGCCGCCCTCGCCGAGAGGCATCATCAA
CGTAGACACAAACAGTCCTAGAACTGTCTATTGGTTGATGGTACCACGTTCTCCTCGCA
GATTGAGATATCTTCCCTCGGAATCTAGGCCTTAAAGAAGGAAGAGTCAGGCAAGAA
GGTGGTGTGTATTCCGGCCCCGGCTGCAATGAACAGCCGATCAAGCGGAGAAGTGGTTCC
ATCGGATCTATGGGCTTGGCGTAAATACGGTCAAAAACCTATCAAAGGCTCTCCTTTTCC
AAGGGGTTATTATAGATGCAGCAGCTCAAAGGTTGTTGAGCAAGAAAGCAAGTCGAAAG
AAGCCGAACCGATCCAAACATGTTGGTGATTACATATACCTCCGAACATAACCATCCTTG
GCCCATCCAACGCAACGCTCTCGCCGGCTCCACACGCTCCTCCACCTCCTCCTCATCTAA
CCCTAATCCTTCCAAACCTCAACCGCAAACGTAACCTCCTCATCCATTGGCTCCCAAAA
CACCATCTACTTGCTTCTCTCCACCACTCCTCCTCCTACCTCTCATCCTCCGCCATCAA
AGATGAACGAGGGGACGATATGGAGTTGGAAAACGTAGATGATGATGATAACCAGAT
TGCTCCATACAGACCGGAGCTTCATGATCATCAGCACCAACCAGATGATTCTTTGCAGA
TCTTGAAGAGCTAGAAGGAGATTCTTAAGCATGTTGCTTTCTCATGGCTGTGGCGGCGA
CGGGAAGGATAAAACGACCGCGTCCGATGGGATCAGCAATTTCTTCGGGTGGTTCGGGAGA
TAATAATTATAATAATTACGACGACCAAGACTCAAGGTCGTTATAGTATAGTGTAAATTA
CAGGTAACAAATTATATTAATTAAGTTGAGCTTGTGAAAATGAAGATCATATGGTCTG
GTCAGGTTGGGGGC

>G182 Amino Acid Sequence (conserved domain in AA coordinates: 217-276)
MCSVSELLDMENFQGLTDVVRGIGGHVLSPETPPSNIWPLPLSHPTSPSDLNINPFGD
PFVSMDDPLLQELNSITNSGYFSTVGDNNNNIHNNGFLVPKVFEEDHIKSQCSIFPRIR
ISHSNIHDSPPCNSPAMSAHVVAASPRGIINVDNTPRNCLLDVDTFFSSQIQIS
SPRNLGLKRRKSQAKKVVCIAPAPAMNSRSSGEVVPDLWAWRKYGQKPIKGSPPFRGYY
RCSSSKGCSARKQVERSRTDPNMLVITYTSEHNHPWPIQRNALAGSTRSSSTSSSSNPNS
KPSTANVNSSSIGSQNTIYLPSTTPPTLSSSAIKDERGDDMELENVDDDDNQIAPYR
PELHDHQHQPDDFFADLEELEGDLSMLLSHGCGDGKDKTTASDGISNFFGWSGDNVYN
NYDDQDSRSL*

>G1835 (1..969)
ATGATTGGAACAAGCTTCCCCGAGGATCTTGATTGTGGCAACTTCTTTGACAACATGGAT
GATCTCATGACTTTCCCGGTGGAGATATCGATGTCGGTTTCGGCATAGGTGACTCCGAC

TCTTTCCCTACCATCTGGACCACTCATCACGACACGTGGCCTGCCGCTTCTGATCCTCTC
TTCTCTTCCAACACCAACTCTGATTCATCACCTGAGCTCTATGTTCCGTTTGAGGACATT
GTTAAGGTGGAAAGACCTCCAAGCTTTGTAGAGGAAACATTGGTTGAGAAGAAGGAAGAT
TCGTTTTTCGACAAACACTGATTCATCATCTTCTCATAGCCAATTGAGGAGCTCAAGTCCA
GTGTCGGTCTCGAGAGCAGCTCCTCCTCGTCTCAAACCACCAACACAACCTCCCTTGT
CTCCCTGGAAAGCAGGTCCTCCACGCACAAAACGCCCTCGTCCACCTGTCCAGGATAAA
GATAGAGTCAAAGACAATGTGTGCGGTGGTGACTCGCGCTCATCATTTAGAATACCGAAA
CAGTTTCTCTCTGATCACAACAAGATGATCAACAAGAAGAAGAAGAAGGCCAAGATT
ACTTCTTCTCTTCTTCTCGTCCGGGATTGATCTTGAAGTCAATGGAACAACGTCGATTCTG
TATTCCTTCAGAGCAATATCCGCTTAGGAAATGTATGCACTGTGAGGTACCAAGACTCCA
CAGTGGAGGCTTGGTCCAATGGGTCCAAAGACACTTTGCAATGCGTGGGTGTACGTTAC
AAATCAGGGAGGCTTTTCCCGGAGTACCGTCCAGCTGCTAGTCCAACATTTACTCCAGCT
CTTCACTCAAACACACAAGAAAGTGGCTGAAATGAGAAACAAGAGATGCAGTGATGGT
AGCTACATAACCGAAGAGAATGATCTGCAAGGGCTGATTCCGAACAATGCCTACATTGGC
GTAGACTAA

>G1835 Amino Acid Sequence (domain in AA coordinates: 224-296)
MIGTSFPEDLDCCNFFDNMDLMDPFGGDIIDVGFIGDSDFSPTIWTTHHDTWPAASDPL
FSSNTNSDSSPELYVPFEDIVKVERPPSFVEETLVEKKEDSFSTNTDSSSSHSQFRSSSP
VSVLESSSSSSQTNTTSLVLPKGKGRPRTRPRPPVQDKRVKDNVCGGDSRLIIRIPK
QFLSDHNKMKKKKKKAKITSSSSSGIDLEVNGNNVDSYSSEQYPLRCKMHCEVTKTP
QWRLGPMGPKTLCNACGVRYKSGRLFPYRPAASPTFTPALHSNSHKVEMRNKRCSDG
SYITEENDLQGLIPNNAYIGVD*

>G1836 (47..610)
ATAACAAGCCTAGAACACTAGAACTTCAAAAAAGAAAAAATCTTATGGAGAACAACAA
CGGCAACAACAGCTGCCACCGAAAGGTAACGAGCAACTGAAGAGTTTCTGGTCAAAGA
GATGGAAGGTAAGTATGATTTCAAAAATCACGACCTTCTATAACTCGTATCAAGAAGAT
TATGAAGTATGATCCGGATGTGACTATGATAGCTAGTGAGGCTCCAATCCTCCTCTCGAA
AGCATGTGAGATGTTATCATGGATCTCACGATGCGTTCGTGGCTCCATGCTCAGGAAAG
CAAACGAGTCACGCTACAGAAATCTAATGTGCGATGCCGCGAGTGGCTCAAACCTGTTATCTT
TGATTTCTTGCTTGATGATGACATTGAGGTAAAGAGAGAGTCTGTTGCCGCCGCTGCTGA
TCCTGTGGCCATGCCACCTATGACGATGGAGAGCTGCCCTCAGGAATGGTAATTGGAAC
TCCTGTTTGTGTAGTCTTGGAATCCACCAACCACAACCACAATGCAGGCATGGCCTGG
AGCTTGGACCTCGGTGTCTGGTGAGGAGGAAGAAGCGCGTGGGAAAAAAGGAGGTGACGA
CGGAAACTAATAAGTGAATACGTTTTAGGGTATTTTCAAGGAATATGTAGTAAATAGT
CATGGATC

>G1836 Amino Acid Sequence (domain in AA coordinates: 30-164)
MENNNNGNNQLPPKNEQLKSFWSKEMEGNLDPKNHDLPIRIKKIMKYDPDVTMIASEAP
ILLSKACEMFIMDLTMRSLHAQESKRVTLQKSNVDAVAQAQVIFDFLLDDDDIEVKRESV
AAAADPVAMPPIIDGELPPGMVIGTPVCCSLGIHQPPQMQAWPGAWTSVSGEEEEARGK
KGGDDGN*

>G1838 (132..1628)
TTCCTTGGCATTCTCTTTAGAACTTTCTGTACAAAATGCAAAACCTGAACCTCTAAAGCTA
AAAAAAAAGATTAGAGACTGTAAGTCTTTTATCAGATTTTCAACTAGGAAAAAAGTTAC
AATCTTTTTTTGATGGCTCCTCAATGACGAATGCTTAACGTTTTCTCTGTCACCAATGG
AGATGTTGAATCAACTGATCAGTCTCACTTCTCTTCTTCTTACGACGATTCCTCTACTC
CTTATCTCATCGATAACTTCTATGCTTTCAAAGAAGAAGCTGAGATAGAAGCTGCTGCTG
CTTCAATGGCGGATTCAACAACCTTATCTACTTTTTTTCGATCATTCTCAGACTCAGATTC
CAAAGCTGGAAGATTCCTCGGTGATTCTTTGTCCGTTACTCTGATAACCAACAGAGA
CCCAAGACTCTTCTTCTCTCACTCCATTCTACGATCCACGTACCCGACCCGTTGCCGAAG
GAGTTACAGGGTTCTTCTGATCATCATCAGCCAGATTCAAGACGATAAACTCGGGAC
CAGAAATCTTCGATGACTCAACAACCTTCCAACATCGGTGGTACTCATCTCTCCAGTCAG
TGGTGGAGTCATCAACGACGCGAAGTTAGGGTTTAACGGTGATTGCACCACCACCGGAG
GAGTTTTGTCTCTAGGGGTTAACAACACATCAGATCAACCTTTGAGCTGTAACAATGGCG
AGAGAGGTGGAACAGTAACAAGAAGAAAACAGTTTCTAAGAAGGAACATCAGATGATT
CAAAGAAGAAGATTGTGCAAAACATTGGGACAAAGAACTTCAATTTATCGTGGAGTCACCC
GACATAGATGGACTGGAAGATACGAAGCGCATCTATGGGATAACAGCTGTAGGAGGGAAG
GTCAAGCCAGAAAAGGACGTCAAGTGTACTTAGGTGGATATGACAAGGAAGATAGAGCAG

CTAGAGCCTATGACTTGGCAGCTTTAAATACTGGGGTTCTACTGCTACTACAAATTTTC
CGGTCTCGAGTTATTCAAAAGAACTTGAGGAAATGAATCACATGACCAAGCAAGAGTTTA
TTGCATCTCTTAGGAGGAAAAGTAGCGGTTTTTCGAGAGGAGCTTCAATATATAGAGGTG
TCACAAGGCATCATCAACAAGGTCGCTGGCAAGCAAGAATCGGCCGTGTCGCAGGAAACA
AAGATCTTTACCTCGGAACCTTTGCAACCGAAGAGGAAGCAGCAGAGGCTTATGACATTG
CAGCCATAAAGTTCAGAGGAATCAACGCAGTAACCTTTGAGATGAACAGGTATGACA
TTGAAGCTGTCAATAGATTCTTTACCTGTAGGAGGAGCAGCTGCGAAACGCCACAAAC
TCAAACCTCGCTCTGAATCTCCTTCTTCATCATCCTCTGACCATAACCTCCAACAACAAC
AGTTGCTTCCGTCCTCTTCTCCCTCGGATCAAAACCTAACTCAATCCCATGTGGCATTG
CATTGAGCCTTCAGTTCTCTATTACCACCAGAAGCTTTTCAGCATTATCCTTTGGTCT
CTGACTCTACAATTCAAGCTCCTATGAACCAAGCTGAGTTTCTTGTGGCCTAACAGT
CTTACTAAATCATTGGTTCTGTTCTGCTTAGACTTCTATTACCGCACTAACCGATGAC
CCGAGGCTTATCTTCTGATTCTGGCTATAAGGATGAATCTTCAAGTTCCTTTTAAAC
TGTAAGCTAAGACAGAAGTAGAGGGGAGAAAAGTTGAAGAATCTGAAACTTTTGGGGTCA
ATTTTGTATTAAATGTTTTCTTTGTCAAGGGTGGATTATCGGTTTTATTACTTATTTT
TGAATGTAATCGGCCTATAACGGTATAACTCTGTTTCCATTATGAATATTTTCTCAA
TTGAAAAAAAAAAAAAAAAA

>G1838 Amino Acid Sequence (conserved domain in AA coordinates: 229-305, 330-400)

MAPPMNCLTFLSPMEMLKSTDQSHFSSSYDDSSPYLIDNFYAFKEEAEIEAAAASMA
DSTTLSTFFDHSQTQIPKLEDFLGDSFVRYSDNQETQDSSSLTFYDPRHRTVAEGVTG
FFSDHHQPDFKTINSQPEIFDDSTTSNIGGTHLSSHVESSTAKLGFNGDCTTTGGVLS
LGVNNTSDQPLSCNNGERGGSNSKKKTVSKKETSDDSKKIVETLQRTSIRGVTRHRW
TGRYEAHLWDNSCRREGQARKGRQVYLGGYDKEDRAARAYDLAALKYWGSTATTNFPVSS
YSKELEEMNHMTKQEFIAFLRRKSSGFSRGASIRGVTRHHQQGRWQARIGRVAGNKDLY
LGTFAEEEAAYDIAAIAKFRGINAVTNFEMNRYDIEAVMNSSLPVGGAAAKRHKLKLA
LESPSSSSSDHNLQQQLLPSSSPSDQNPNSIPCGIPFEPVLYYHQNFFQHYPLVSDST
IQAPMNQAEFFLWPNQSY*

>G1843 (51..653)

CAGACATCACAATCAAATTAGGTCAGAAGAATTAGTCGGAGAAAACAGCCATGGGAAGAA
GAAAAGTAGAGATCAACGAATTGAGAACAAAAGCTCTCGACAAGTTACTTTCTGTAAAC
GACGAAATGGTCTCATGAGAGAAAGCTCGTCAACTCTCAATCTTTGTGAATCCTCCGTCG
CTCTTATCATCATCTCTGCCACCGAAGACTCTACAGCTTCTCTCAGGTGATAGCATGG
CCAAGATCCTCAGTCGTTATGAATTAGAACAGGCTGATGATCTTAAACCTTGGATCTAG
AAGAAAAACTCTTAATTATCTTTGCGACAAGGAGTTGCTAGAAACAATCCAATGCAAGA
TTGAAGAAGCGAAAAGCGATAATGTAAGTATAGATTGTCTAAAGTCCCTGGAAGAGCAGC
TCAAGACTGCTCTGTCTGTAAGTAGAGCTAGGAAGACAGAACTAATGATGGAGCTTGTGA
AGACCCATCAAGAGAAGGAGAAGCTGCTGAGAGAGGAGAACCAGAGTTTGAATAACCAGC
TTATAAAGATGGGGAAGATGAAGAAGTCTGTGGAAGCAGAGGATGCAAGAGCAATGTCAC
CGGAAAGTAGCTCTGACAACAAGCCACCGGAGACTCTCCTGCTTCTCAAGTAACCACCAT
CACCAACGACTGATTGCAAAAAATAAAATTGTAATAATTATGATTTGTAGTTCATAAGGA
AAGCTACATACTGTATGTTAAAAATCCTTCTTCCCCTGCTACGGAAGAGTCATCCAA
GGAGATGCATCAAATAAAGTAATTGATTTTATTGTTA

>G1843 Amino Acid Sequence (domain in AA coordinates: 2-57)

MGRRKVEIKRIENKSSRQVTFCKRRNGLMEKARQLSILCESSVALIIISATGRLYSFSSG
DSMAKILSRYELEQADDLKTLDLEEKTLNLYSHKELLETIQCKIEBAKSDNVSIDCLKSL
EEQLKTALSVTRARKTELMELVKTHQEKEKLLREENQSLTNQLIKMGKMKKSVEAEDAR
AMSPSSSDNKPPETLLLLK*

>G1853 (1..1860)

ATGAGAGGTTCTTGGTACAAGAGTGTTTCTCTGTTTTTGGTCTCAGACCACGGATCAGA
GGGTTGTTATTCTTCATTGTTGGTGTGTGGCTCTAGTTACTATTTTAGCACCATTGACA
TCTAATTCGTATGATTCTTCGTCAAGTTCGACACTTGTGCCGAACATTTATAGTAACTAT
AGGAGGATAAAGGAGCAAGCTGCTGTTGATTATCTTGATCTGAGGTCTCTTTCTTTAGGG
GCTAGTTTTAAAGAGTTTTCTTTTGTGGTAAAGAAAGAGAAAGTTATGTGCCTTGTAT
AACATAACTGGGAATTTGCTTGCTGGGCTTCAAGAGGGTGAGGAGTTAGATCGACATTGC
GAGTTTGAAAGAGAGAAGGAAAGATGTGTAGTTCGTCTCCGAGAGATTATAAAATACCA
CTTAGGTGGCCACTTGGTAGAGATATCATATGGAGTGGGAACGTGAAGATTACCAAAGAC
CAGTTTCTTTCTTCAGGAACGTGACAACGAGGTTAATGTTGCTTGAAGAGAATCAAATA

ACCTTTCACTCGGAGGACGGCCTGGTCTTTGATGGGGTCAAAGACTATGCTCGTCAAAT
GCTGAGATGATAGTTTATAGGAAGTGATACTGAATTTGCTCAAGCGGGTGACGGACTGTG
TTAGACATTGGTTGCGGATTTGGTAGCTTTGGTGCTCATTTAGTGTCTTTGAAGCTGATG
CCTATATGTATTGCTGAGTATGAGGCAACTGGGAGCCAAGTTCAGTTAGCTCTAGAGAGA
GGCCTTCCTGCAATGATTGGCAATTTCTTTTCAAACAGCTTCCTTATCCAGCACTGTCT
TTTGATATGGTCCATTGTGCTCAATGTGGCACTACTTGGGATATCAAAGATGCAATGCTA
CTTTTGGAAAGTGGATCGTGTCTGAAACCCGGGGGATACTTTGTTTAACTTCTCCACA
AACAAAGCACAGGGAACTTACCAGATACCAAGAAAACGAGCATCTCAACACGGGTGAAT
GAGTTATCTAAGAAAATCTGTTGGAGTCTAACAGCTCAGCAGGATGAGACGTTTCTTTGG
CAGAAAACCTTCTGATTCAAGTTGCTATTCTTCTCGTTCGCAAGCTTCTATACCTCTTTGC
AAAGATGGAGATAGCGTTCCGTATTACCACCCATTGGTTCCATGTATAAGCGGAACCACG
AGTAAACCGCTGGATTTCTATACAGAACAGGTCTGCTGTTGCAGGAACAACCTCTGCCGGG
CTTGAAATTCATGGTTTAAACCCGGAAGAATTCTTCGAGGATACACAAATATGGAGATCA
GCTCTGAAAACTATTGGTCTTGTCTTACACCTCTAATTTTCTCTGACCATCCGAAGAGA
CCCGGTGATGAGGATCCCTCTCCCGCCTTTCAACATGATACGCAATGTGATGGACATGCAT
GCTCGTTTGGGAAATTTAAATGCCGCTTTACTCGACGAAGGAAAATCTGCTTGGGTAAATG
AACGTCGTCCTCAGTCAATGCACGTAATACTCTTCTCTATCATACTTGATCGTGGTTTCGCC
GGTGTCTACATGACTGGTGTGAACCATTCCTCGACATATCCTCGAACATATGACATGCTT
CATGCCAATGAATTTCTCACACATCTTAGCTCAGAACGATGCAGCCTAATGGACTTGTTT
TTGGAGATGGACCGGATTCTTCGCCCTGAGGGATGGGTGTTCTAAGCGACAAAGTGGGA
GTAATCGAGATGGCTCGACACTTGCAGCTCGAGTGCCTTGGGAAGCAAGAGTCATTGAT
CTTCAAGATGGTAGTGACCAAGACTTCTCGTCTGTCAAACCATTCATCAAAAAATAA
>G1853 Amino Acid Sequence (domain in AA coordinates: entire protein)
MRGSWYKSVSSVFLRPRIRGLLFFIVGVVALVTILAPLTSNSYDSSSSSTLVPNIYSNY
RRIKEQAAVDYLDLRSLSLGASLKEFPFCGKERESYVPCYNITGNLLAGLQEGEELDRHC
EFEREKERCVVRPRDYKIPLRWPLGRDIIWSGNVKITKDQFLSSGTVTTRLMLLEENQI
TFHSEDLGVFDGVKDYARQIAEMIGLGSDEFAQAGVRTVLDIGCCGFSFGAHLVSLKLM
PICIAEYBATGSQVQLALERGLPAMIGNFFSKQLPYPALSFDMVHCAQCQTWDIKDAML
LLEVDRLVKPGGYFVLTSPTNKAQGNLPDTKKTSISTRVNELSKKICWSLTAQQDETFW
QKTS DSSCYSSRSQASIP LCKDGSVPYHPLVPCISGTTSKRWISIQNRSAVAGTTSAG
LEIHGLKP EEFEDTQI WRSALKNYWSLLTPLIFSDHPKRPGEDEPLPPFNMRNVMDMH
ARFGNLNAALLDEGKSAWVMNVVPVNARNTLP IILDRGFAGVLHDWCEPFPTYPRYDML
HANELLTHLSSERCSLMDLFLEMDRILRPEGWVVLSDKVGVIEMARALAAVRVWEARVID
LQDGS DQRL LVCQKPF IKK*
>G1855 (1..1902)
ATGGCGAAAGAGAACAGTGGTCATCATCACCAAACAGAAGCAAGAAAGAAAGAACTAACT
TTGATTCTTGGTGTAAGTGGACTCTGCATTTTGTCTATGTTTATAGGTGCATGGCAAGCC
AATACCGTCCCATTCTATCTCGAAGCTCGGATGCGAGACGCAATCAAACCTTCTTCG
TCCTCTTCTCTTCTCATCTTCAGAGTCAGCTGAAGTAGATTCAAAGCCATAATCAG
ATTGAGTTAAAGGAAACAAACCAACCATTAAGTACTTTGAACCATGTGAATTATCTCTC
AGTGAGTACACTCCTTGTGAAGACCGACAAAGAGGAAGAAGATTGATAGGAACATGATG
AAATATAGAGAAAGACATTGTCTGTAAAAGATGAGCTTCTTTATTGTTTGATTCTCCT
CCACCAAAC TACAAGATTCCATTTAAATGGCCACAAAGTAGAGACTATGCTTGGTATGAC
AATATCCCTCACAAGGAACCTTAGTGTGAGAAAGCAGTTCAAACTGGATTCAAGTTGAA
GGTGACCGCTTTTAGATTCCCTGGTGGTGGTACTATGTTTCTCGTGGAGCTGATGCTTAT
ATCGATGATATTGCTAGGCTTATTCTCTTACTGATGGTGAATCAGAACAGCTATTGAC
ACTGGATGTGGTGTGCAAGTTTGGTGCTTACCTCTTGAAGAGAGACATTATGGCTGTG
TCTTTTGTCTCAAGAGACTCATGAAGCTCAGGTACAGTTTGCTTTAGAACCGCGAGTT
CCTGCGATAATCGGATATGGGATCAAGAAGACTTCTTATCCAGCTAGAGCTTTTGAT
CTTGCTCATTGTTCTCGTTGTGTTGATCCCTTGGTTTAAAAATGATGGTTTGTACCTTATG
GAGGTCGACCGGTTTAAAGACCGGGCGGTTACTGGATCCTCTCGGGACCACCGATTAAAC
TGGAAACAGTACTGGAGAGGGTGGGAGAGAACAGAGGAGGATTTGAAGAAAGAGCAAGAT
TCAATAGAAGATGTAGCAAAGAGTCTTTGCTGGAAGAAAGTAACTGAAAAAGGTGACTTA
TCAATTTGGCAAAAGCCTCTCAATCACATTGAGTGTA AAAAGCTCAAACAAAACAATAAG
TCACCTCCGATATGCAGATCAGATAACGCGGATTCCGCTTGGTACAAAGACTTGGAACT
TGATAACACCAATTACCAGAAACAAACAATCCAGATGATTACAGAGCGGTGCACTCGAG
GATTGGCCAGACCGAGCATTCGCGGTACCTCCAAGAATCATCAGAGGAAC TATACCAGAA

ATGAACGCGGAGAAATTTAGAGAAGACAACGAGGTTTGGAAAGAGAGAATAGCACATTAC
AAGAAGATAGTCCCTGAGCTTTACATGGAAGATTACGGAACATTATGGACATGAACGCT
TTTCTCGGCGGATTTCGCTGCTTCCATGCTGAAATATCCCTCATGGGTCATGAACGTTGTC
CCGGTCGATGCAGAGAAACAAACGTTAGGTGTGATCTACGAACGTGGATTGATAGGGACG
TATCAAGATTGGTGTGAAGGATTCTCAACGTATCCAAGAACTTATGATATGATTTCATGCA
GGAGGATTGTTTCAGCTTATACGAACATAGGTGTGATTTGACGTTGATATTGTTGGAGATG
GATCGAATTTTTCAGACCAGAAGGAACAGTTGTGTTGAGAGATAATGTGGAGACGTTGAAT
AAGGTAGAGAAGATAGTGAAGGGAATGAAGTGAAGAGTCAAATTGTTGATCATGAGAAA
GGTCCTTTTAATCCTGAGAAGATTCTTGTGCTGTTAAACTTATTGGACTGGTCAACCT
TCTGACAAGAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA

>G1855 Amino Acid Sequence (domain in AA coordinates: entire protein)

MAKENS GHHHQTEARRKKLTLLILGVSLCILFYVLGAWQANTVPSSISKLG CETQSNPSS
SSSSSSSSSES AELDFKSHNQIELKETNQTIKYFEPCELSLSEYTPCEDRQGRFRFDRNMM
KYRERHCPVKDELLYCLIPPPNYKIPFKWPQSRDYAWYDNIPHKELSV EKAVQNW IQVE
GDRFRFPGGGTMFPRGADAYIDDIARLIP L TDGGIRTAIDTGCGVASFGAYLLKRDIMAV
SFAPRDTHEAQVQFALERGVP AIIIGIMGSRRLPYPARAFDLAHSRCLIPWFKNDGLYLM
EVDRLVRPGGYWILSGPPINWKQYWRGWERTBEDLKEQDSIEDVAKSLCWKKVTEKGD L
SIWQPLNHIIECKLKLQNKSPPICSSDNADSAWYKDLETCTITPLPETNNPDD SAGGALE
DWPDRAF AVPPRIIRGTIPEMNAEK FREDNEVWKERIAHYKKIVPELSHGRFRNIMDMNA
FLGGFAASMLKYP SWVMNVVPDAEKQTLGVIYERGLIGTYQDWCEGFSTYPR TYDMIHA
GGLFSLYEHRCDLTLILLEM DRILRPEGTVVLRDNVETLNKVEKIVKGMKWK SQIVDHEK
GPFNPEKILVAVKTYWTGQPSDKNNNNNNNNNN*

>G187 (118..1074)

TAGACCTCTTAGGAAAAAACCTAAAAACCTAATCCCCAACCTAAAAGGCTTATCTCAT
CTCTTCTCTTTGTCTTTACTCTTTTTTTTACCTCTCTCTCATTGTTCTTCAACATG
TCTAATGAAACAGAGATCTCTACAAC TACCAATACCCCTTCATCGTTTTCGTTGCACGAA
ATGATGAATCTGCCTACTTCAAATCCATCTTCTTATGGAACCTCCCATCACAAAACGGT
TTTAATCCATCTACTTATTCCTTACCGATTGTCTCAAAGTTCTCCAGCAGCGTATGAA
TCTCTACTTCAGAAAAC TTTTGGTCTTTCTCCCTCTTCTCAGAGGTTTTCAATTCTTCG
ATCGATCAAGAACC GAACCGTGATGTTACTAATGACGTAATCAATGGTGGTGCATGCAAC
GAGATTCGAGTAAAGGTTTCTCCTTCTAATTCTTCTCTAGTGAGGCTGATCACCCCGGT
GAAGATTCGGTAAAGCGCGGAGGAAACGAGAGTTAGTGGTGAAGAAGATCAAATTTCC
AAAAAAGTTGGGAAAACGAAAAGACTGAGGTGAAGAAACAAAGAGAGCCACGAGTCTCG
TTTATGACTAAAAGTGAAGTTGATCATCTTGAAGATGTTATAGATGGAGAAAATACGGC
CAAAAGGCTGTAAAAATAGCCCTTATCCAAGGAGTTACTATAGATGTACAACACAAAAG
TGCAACGTGAAGAAACGAGTGGAGAGATCGTTCCAAGATCCAACGGTTGTGATTACAAC T
TACGAGGGTCAACACAACCAACCCGATTCCGACTAATCTTCGAGGAAGTTCTGCCGCGGCT
GCTATGTTCTCCGACACCTCATGACTCCAAGAAGCTTGCACATGATATGTTTAGGACG
GCAGCTTATACTAACGGCGGTTCTGTGCGCGCGGCTTGGATTATGGATATGGACAAAGT
GGTTATGGTAGTGTGAATTCAAACCC TAGTTCTCACCAAGTGTATCATCAAGGGGGTGAG
TATGAGCTCTTGAGGGAGATTTTTCCTTCAATTTCTTAAAGCAAGAGCCTTGATCGATC
ATTGTTATAACTACATATATTATATATATGAGAGAGAGAGGTAGAGAAAAA

>G187 Amino Acid Sequence (domain in AA coordinates: 172-228)

MSNETRDLYNQYPSSFSLHEMMNLPTSNPSSYGNLPSQNGFNPSTYSFTDCLQSSPAAY
ESLLQKTFFGLSPSSSEVFNSSIDQEPNRDVTNDVINGGACNETETRVSPSNSSSSEADHP
GEDSGKSRRKREL VGEEQISKVVGKTKKTEVKKQREPRVSFMTKSEVDHLEDGYRWRKY
GQKAVKNSPYPRSYRCTTQKCNVKKRVERSFDPTVVITTYEGQHNHPIPTNLRGSSAA
AAMFSADLMTPRSFAHDMFRTAAYTNGGSVAAALDYGYSQSGYGSVNSNPSSHQVYHQGG
EYELLREIFPSIFFKQEP*

>G1881 (1..519)

ATGCGAATTTTGTGTGATGCTTGTGAGAGCGCCGCCGCTATCGTCTTTTGGCGCCCGAC
GAAGCTGCCCTCTGTTGCTCTGCGACGAAAAAGTTCATAAGTGCAACAAGCTGGCTAGT
CGGCATCTTCGTGTAGCTTAGCTGATCCGAGTAATGCACCAAGCTGTGACATATGCGAA
AATGCACCCGATCTTTTTACTGTGAGATAGATGGTAGTTCCCTTTGTCTACAATGTGAT
ATGGTGGTACATGTTGGTGGGAAGAGAACACATAGGCGGTTTCTATTACTGAGACAGAGA
ATTGAGTTTCCAGGCGATAAGCCTAATCATGCTGACCAACTGGGACTACGGTGTCAAAG
GCTTCCTCTGGTCTGGTCAAGAATCAAATGGGAATGGTGATCATGATCATAATATGATC

GATCTTAACTCCAATCCTCAAAGAGTACACGAGCCTGGATCACATAACCAAGAGGAGGGT
ATTGATGTAAATAACGCAAACAATCAGGAGCATGAATAG
>G1881 Amino Acid Sequence (domain in AA coordinates:5-28, 56-79)
MRILCDACESAAAIVFCAADEAALCCSCDEKVHKCNKLASRHLRVGLADPSNAPSCDICE
NAPAFYCEIDGSSLCLQCDMVVHVGGKRTHRRFLLLRQRIEFPDGPKNHADQLGLRCQK
ASSGRGQESNGNDHDHNMIDLNSNPQVRVHEPGSHNQEEGIDVNNANNHEHE*
>G1882 (1..1200)
ATGGTTTTTCTTCATTTCTTACTTATCCTGATCATTTCATCAAACCTGGCAACAACAACAT
CAACCAATCACAAACCACCGTTGGATTACGGGAAATAACATCAACCAACAGTTTCTTCCT
CACCATCCCCCTCCACCGCAACAGCAACAAACGCCTCCGCAGCTTCACCACAACAACGGT
AACGGCGGAGTCGTGTTCCCGTGGACCTGGCGGGTTAATCCGACCAGGTTTCGATGGCG
GAAAGAGCAAGGCTAGCCAACATACCATTACCTGAAACAGCCTTGAAGTGTCCAAGATGT
GACTCAACTAACACCAAATTCGTGTTACTTCAACAACTACAGTCTCACTCAACCTCGCCAC
TTCTGCAAAGCATGCCGTGTTACTGGACACGTGGCGGTGCTTAAGGAGCGTTCCCGTC
GGTGGCGGTTGCCGTAGAAACAAAAGAACAAAAACAGCAGCGGTGGAGGTGGCGGTAGC
ACCACTAGCGGTAAACAGCAAGTCACAAGACAGCGCCACGAGCAACGACCAATACCACCAC
CGAGCCATGGCTAACAAATCAGATGGGACCACCTTCTTCGTTCATCGTCTCTAAGCTCGTTG
CTGTCTTCTTACAACGCAGGGTTAATCCCCGACATGATCATAACAGCAATAACAACAAC
ATACTTGGACTTGGATCATCTTTGCCTCCTCTAAGCTTATGCCTCCTTTAGACTTCACA
GACAACTTCACCTTACAATACCGTGCCGTTTCAGCTCCTTCTTATCATATAGCGGTGGA
AGCAGTGGAGGAGCGGCGGCTCTTTAAACGGTTTGGACAGTGGAGATTCCCGGCAACA
AACCAACTTCTTTAGGCGGTTTAGACCCGTTTGATCAACAACATCAAATGGAGCAGCAG
AATCCAGGTTACGGATTGGTTACCGGGTGGGTTCAGTATCGACCTAAGAACATTTTCCAT
AACCTTATCTCCTTCTTCTCGTCTGCTTCATCAGCTATGGTTACAGCCACCGCGTCGCAA
TTAGCTTCAGTGAAATGGAAGATAGTAACAATCAGCTCAACTTGTCTAGACAACCTTTT
GGAGACGAACAACAGCTCTGGAATATTCATGGCGCTGCTGCAGCATCCACCGCAGCTGCA
ACAAGTTCGTGGAGTGAAGTCTCTAATAATTTAGTTCTTCTTCTACTAGCAATATATAA
>G1882 Amino Acid Sequence (domain in AA coordinates:97-125)
MVFSFPTYPDHSSNWQQHQPIITTVGFTGNINQQFLPHHPLPPQQQQTTPQLHHNNG
NGGVAVPGGPGGLIRPGSMAERARLANIPLPETALKCPRCDSTNTKFCYFNYSLTQPRH
FCKACRRYWTRGGALRSVPVGGGCRNRKTKNSSGGGGSTSSGNSKSQDSATSNDQYHH
RAMANNQMGPSSSSSSLLSSYNAGLIPGHDHNSNNNNIILGLSSLPPLKLMPPPLDFT
DNFTLQYGAVSAPSYHIGGGSSGAAALLNGFDQWRFPATNQLPLGGLDPFDQQHQMEQQ
NPGYGLVTGSGQYRPKNIFHNLISSSSSASSAMVTATASQLASVKMEDSNNQLNLSRQLF
GDEQQLWNIHGAAAAATAATSSWSEVSNFSSSSTSNI*
>G1883 (1..1110)
ATGGACGCTACGAAGTGGACACAGGGTTTTCAAGAAATGATGAACGTTAAACCAATGGAG
CAGATCATGATTCCTAATAACAACACACATCAACCAACACCACATCCAATGCAAGGCCA
AACACCAATTCTCACATCTAACGGCTCTCAACTGCTGGAGCAACCGTCTCCGGCGTAAGC
AACACAATAACAATACGGCGGTTGTGGCGGAGAGGAAAGCAAGACCACAAGAGAACTA
AATGTGCAAGATGCAACTCAACCAACACAAAGTTTTGTTACTACAACAATATAGTCTC
ACACAACCAAGATACTTCTGCAAAGGTTGTGGAAGGTATTGGACCGAAGGTGGATCTCTT
AGGAATGTTCTGTGGGAGGAAGCTCAAGAAAGAACAAAGAGATCATCTTCATCTTCTTCA
TCAAACATCCTTCAGACAATACCATCTTCACCTCCAGATCTAAACCCGCCAATACTCTTC
TCAAACCAATCCATAATAAATCGAAAGGGTCATCACAAGATCTCAACTTGTGTCTTTC
CCAGTCATGCAAGATCAACATCATCATCATGTCCATATGTCTCAGTTTCTTCAGATGCCT
AAGATGGAGGGAAATGGTAACATAACTCATCAGCAGCAGCCTTCATCATCTTCTTCTGTC
TATGGTTCCTCGTCCTCTCTGTTTCAGCTCTTGAACTTTTAAGAACCGGAGTTAATGTT
TCTTCAAGATCAGGATTAACTCATCGTTCATGCCTTCCGGTTCAATGATGGATTCAAAC
ACTGTGCTTTACACTTCTTCAGGGTTTCCACAATGGTGGATTACAAGCCAAGTAATCTC
TCCTTCTCTACCGATCATCAAGGGCTTGGACACAATAGCAACAATAGGTCTGAAGCTCTT
CATAGTGATCATCACCAACAAGGTAGAGTTTTGTTTCCATTTGGGGATCAAATGAAGGAG
CTTTCATCAAGCATAACACAAGAAGTTGATCATGATGATAATCAACAACAGAAGAGTCAT
GGAAATAATAATAATAATAACTCAAGCCCTAATAATGGATATTGGAGTGGGATGTTT
AGTACTACAGGAGGAGGATCTTCATGGTGA
>G1883 Amino Acid Sequence (domain in aa coordinates: 82-124)
MDATKWTQGFQEMNVKPMEQIMIPNNNTHQPNTTSNARPNTILTSNGVSTAGATVSGVS

NNNNNTAVVAERKARPQEKLNCPRCNSTNTKFCYNNYSLTQPRYFCKGCRRYWTEGGSL
RNPVPGGSSSRKNKRSSSSSSSSNILOTIPSSLPDLNPPILFSNQIHNKSKGSSQDLNLLSF
PVMQDQHVVHMSQFLQMPKMEGNGNITHQQQPSSSSSVYSSSSPVSALELLRTGVNV
SSRSGINSSFMPSGSMDSNTVLYTSSGFPTMVDYKPSNLSFSTDHQGLGHNSNNRSEAL
HSDHHQQGRVLFPPFDQMKELSSSITQEVDDDDNQQKSHGNNNNNNSSPNNGYWSGMF
STTGGGSSW*

>G1884 (1..741)

ATGATGACGTCATCCCATCAGAGCAACACCACCGGCTTTAAACCGCGGCGGATCAAGACG
ACGGCGAAGCCACCACGTCAGATCAATAACAAAGAACCATCTCCGGCGACGCAGCCGGTG
CTCAAGTGTCCGAGATGTGATTCAAGTCAACACCAAATTCTGTACTACAACAACACTACAGC
TTGTCTCAGCCACGTCCTACTACTGCAAGAACTGTCTGCTTACTGGACACGTGGCGCGGCC
CTCCGTAACGTTCCCATCGGTGGCTCCACTCGAAACAAGAACAAGCCTTGCAGCCTCCAA
GTCATCTCTTCTCTCTTTGTTCTCGAACGGGACGTCATCGGCGTCTCTGTGAGCTTGTA
AGAAACCATCCATCGACGGCAATGATGATGATGAGTTCTGGTGGATTCTCCGGCTATATG
TTTCCGTGGATCCTAACTTCAACCTTGCCTCGTCTTCTATCGAGTCTTTGAGTTCTTTT
AACCAAGATTGACCCAGAAGCTTCAGCAACAAAGACTCGTCACTTCCATGTTTCTCCAA
GATTCTCTTCCGGTTAACGAGAAAACGGTTATGTTTCAGAACGTAGAGTTGATTCTCTCT
TCGACGGTGACGACGATTGGGTTTCGATAGGTTCCGCACTGGAGGAGGTGCAACAAGT
GGCAATCATGAAGATAATGATGATGATGGGGAGGGTAATTGGGAAATTGGTTCCATAATGCT
AATAATAATGCTCTGCTCTAA

>G1884 Amino Acid Sequence (domain in AA coordinates:43-71)

MMTSSHQNTTGFKPRRIKTTAKPPRQINNKEPSPATQPVLCPCDVSNTKFCYNNYS
LSQPRHYCKNCRRYWRGGALRNVPPIGGSTRNKNKPCSLQVISSPPLFSNGTSSASREL
RNHPSTAMMMSSGGFSGYMFPLDPNFNLAASSIESLSFNQDLHQKLQQQLVTSMLFQ
DLSLPVNEKTVMFQNVELIPPSTVTTDWFDRFATGGGATSGNHEDNDDEGNLGNWFHNA
NNNALL*

>G1891 (1..750)

ATGGATAACTTGAATGTTTTTCGCAAATGAAGACAATCAAGTGAATGATGTGAAGCCCCA
CCACCACCACCTCGAGTGTGTGCAAGGTGTGATTCTGATAATACTAAATTTTGTTATTAC
AACAACACTACTGTGAGTTTCAGCCACGATACTTCTGCAAGAACTGTCGTAGATACTGGACT
CATGGTGGGGCTTTAAGAAACATACCAATTGGTGGAGTAGTCGTGCCAAACGGGCAAGG
GTAAATCAACCTTCGGTTGCTCGGATGGTTTCTGTTGAGACCAACGAGGTAACAATCAA
CTTTCTCTAATGTTCAAGAAAACGTTTCATCTTGTGGATCTTTTGGTGCTTCATCTTCA
TCTTCTGTGGTGCTGTGGGAACCTTTTGGTCTTTGTATGATATTCATGGTGGTATG
GTAACAAATTTGCATCCAACCTCGAAGTGTTCGACCAATCATCGCTTAGCTTTCCATGAT
GGATCATTTGAGCAAGACTATTACGATGTTGGGTCCGATAATCTTTGGTCAACCAACAA
GTTGGTGGCTACGGTTATCACATGAATCCAGTGGATCAATTCAAGTGAACACAGAGCTTC
AACAACACTATGAACATGAATTATAATAACGATAGCACTAGTGAAGTAGCAGAGGATCT
GACATGAATGTGAACCATGATAACAAGAAGATCAGATACCGCAACTCTGTGATTATGCAT
CCTTGTCATCTGGAGAAGGATGGTCTTGA

>G1891 Amino Acid Sequence (domain in aa coordinates: 27-69)

MDNLNVFANEDNQVNDVKPPPPPPRVRCARCDSDNTKFCYNNYCEFPQRYFCKNCRRYWT
HGGALRNPIPGGSSRAKRARNVQPSVARMVSVETQRGNNQPFQSNVQENVHLVGSFGASS
SSVGAVGNLFGSLYDIHGMVITNLHPTRTRVRPNHRLAFHDGSFEQDYDVGSNDLLVNQ
VGGYGYHMPVDQFKWNQSFNNTMNMNYNDSTSGSSRSGSDMNVNHDNKKIRYRNSVIMH
PCHLEKDG*

>G1896 (1..951)

ATGTCCTCCCATAGCAATCTCCCTCTCCCAAACAGTTCCTAAACCAGATCACCGTATC
TCCGGTACATCCCAAACCAAGAAACCACCGTCTTCTCCGTAGCTCAAGACCAACAAAC
CTAAATGCCCTCGTTGCAACTCTCCAAACACAAAGTTCTGTTACTACAACTACAGT
CTCTCTCAACCTCGTCACTTCTGCAATCTTGTCCCGTTACTGGACACGTGGCGGTGCT
CTAAGAAACGTCCCCATCGGTGGTGGTTGCCGGAACCAAAAATCTATCAACCTAAT
TCCTCCATGAACAACTTCTTCTTCTTCTCTCTCAGAGTTCTTCTCATCAATCATG
GAAGATTCATCCAAATCTTCCCTCCTCCGACAAATGATTTTTCAGCTGGCCGGATTA
TCTCTCAACAAAATGAACGATCTTCAACTTTTGAATAACCAAGAAGTTCTTGATCTTAGG
CCCATGATGTCCTCGGGCCGAGAAAACACACCCGTTGATGTCGGGTGGGTTTATCCCTA
ATGGGTTTTGGAGATTTCAACAACAACCATCACCGACGGGGTTCACAACCCCGGAGCA

AGCGACGGAAACTTAGCTTCTTCTATAGAGACTTTGAGTTGTTTAAACCAAGATTTACAC
TGGAGGCTTCAGCAACAGAGGATGGCGATGCTTTTTGGTAATTCCTAAGGAAGAACTGTT
GTCGTCGAGAGGCCACAACCTATTCTTTATCGGAATCTTGAGATCGTAAACTCATCATCG
CCGTCGTCGCCGACGAAGAAAGGAGATAATCAGACAGAGTGGTATTTTGGTAATAACAGT
GATAATGAAGGAGTGATTAGTAATAATGCTAATACAGGAGGAGGAGGAAGTGAATGGAAC
AATGGAATTCAAGCTTGGACTGATCTTAATCATTATAATGCATTGCCTTGA

>G1896 Amino Acid Sequence (domain in aa coordinates: 43-85)
MSSHTNLPSPKPVKPDHRISGTSQTKKPPSSSVAQDQQLKPCRCNSPNTKFCYNNYS
LSQPRHFCKSCRRYWRGALRNVP.IGGGCRKTKKSIKPNSSMNTLPSSSSSQRFSSIM
EDSSKFFPPPTTMDFLAGLSLNKMNDLQLLNNQEVLDLRPMSSGRENTPVVDVGSGLSL
MGFGDFNNHNSPTGFTTAGASDGNLASSIETLSCLNQLDHLWRLLQQRMAMLFNGSKEETV
VVERPQPILYRNLEIVNSSSPSPTKKGDNQTEWYFGNNSDNEGVISNNANTGGGSEWN
NGIQAWTDLNHYNALP*

>G1898 (1..630)
ATGCCGTCGGAACCAACCAACCCGACCCACCAGAGTTCAGCCCTCAACGGCGGCTTAC
CCACCGCCAAATCTGGCTGAGCCTCTTCCTTGCTCCTCGCTGCAACTCCACCACCACCAAG
TTCTGTACTACAACAACTATAACCTCGCTCAGCCTCGCTACTACTGCAAATCTTGCCGC
CGTTACTGGACTCAAGTGGTACACTCCGTGACGTCCCGCTCGGTGGTGGAACTCGTCGA
AGCTCCTCAAAACGTACCGGTTCTTTCTCCACCCTGCCACCTCTCTTCTCCTCTTCT
TCCGTCAATCACCACCACGACACAAGAACCAGCCAGACTGAAGCGAGTCAAACCTAAGGTT
ACTAATTTAATTTAGTTCATGGAAGCTTTGCTTCTCTGTTAGGTTTAGGAAGTGGAAT
GGTGGGTTGGATTACGGGTTTGGGTACGGGTACGGGCTTGAGGAGATGAGTATTGGGTAT
CTTGGAGATTCTTCCGTAGGAGAGATTCCGGTGGTTGATGGTTGTGGTGGTGACACGTGG
CAGATTGGGGAGATTGAAGGTAAAGTGGAGGAGACAGTTTGATATGGCCTGGTCTTGAG
ATCTCAATGCAAACCAACGATGTTAAGTGA

>G1898 Amino Acid Sequence (domain in AA coordinates: 31-59)
MPSEPNTQTRPVRQSTAAAYPPPPLAEPLPCPRCNSTTKFCYNNYNLAQPRYCKSCR
RYWTQGGTLRDVPGGGTRRSSSKRHSFSTTATSSSSSSSVITTTTQEPATTEASQTKV
TNLISGHGSFASLLGLSGNGGLDYGFYGYGLEEMSIGYLGDSVGEIPVVDGCGGDTW
QIGEIEGKSGGDSLIWPGLEISMQTNDVK*

>G1902 (1..615)
ATGCAGGATCCAGCAGCATATTACCAGACGATGATGGCGAAGCAACAACAACAACA
CCACAGTTTGCAGAGCAAGAACAGTTAAAGTGTCTCGTTGTGACTCACCAAACACTAAA
TTCTGTACTACAACAACATAATCTCTCACAGCCTCGTCACTTTGCAAAAGCTGTCTGT
CGTTACTGGACTAAAGCGCGCTCTCCGTAACGTTCCCGTCGGTGGTGGTTCTCGTAAG
AACGCAACCAACGATCCACTTCTTCTTCTTCTTCTGCTTCTCTTCTCAACAGTAGC
CAAAACAAGAAGACGAAAAACCCGGATCCGGATCCTGATCCAGTAATTCTCAAAAACCG
GATTTGGATCCGACCCGGATGCTTTACGGGTTTCCGATCGGTGACCAAGACGTGAAGGGT
ATGGAGATTGGTGGAGCTTTAGCTCGTTGTTGGCGAATAATATGCAGCTTGGTCTTGGA
GGAGGAGGGATCATGCTTGACGGGTCCGGTTGGGATCATCCGGTATGGGTTTGGGTTTG
AGGAGAACCGAACCGGTAATAATAATAATAAATCCATGGACCGATCTGGCTATGAACAGA
GCGGAGAAAAACTGA

>G1902 Amino Acid Sequence (domain in AA coordinates:31-59)
MQDPAAYYQTMMAKQQQQQPPQFAEQEQLKPCRCDSPTKFCYNNYNLSQPRHFCKSCR
RYWTKGALRNVPVGGGSRKNATKRSTSSSSSASSPSNSSQNKTKNPDPPDPNRSQKP
DLDPTRMLYGFPIGDQDVKMEIGGSFSSLLANNMQLGLGGGIMLDGSGWDHPPMGLGL
RRTEPGNNNNNPWTDLAMNRAEKN*

>G1904 (1..924)
ATGCAAGATATTCTAGATTCTCCATGAACGGAGTTGGTGGTGGGGGAGGAGGAGGGG
AGGTTTTTTCGGTGGAGGAATCGCGCGGAGGAGGTGGTGATCGAAGGATGAGAGCTCAT
CAGAACAATATACTTAACCATCATCAATCTCTCAAGTGTCTCGTTGTAATTTCTTTAAC
ACAAAGTTCTGTACTACAACAATTACAATCTTTCTCAGCCTCGTCACTTTTCGAAGAAC
TGTCGTCGTTACTGGACTAAAGGTGGTGTCTCCGTAACGTTCCCGTCGGAGGTGGTTGC
CGGAAAGCTAAACGTTTGAAACAAAACAGGTTCCGTCGTCGTCATCAGCCGACAAACCA
ACGACGACGCAAGATGATCATCAGTGGAGGAGAAATCGAGTACAGGATCTCACTTAGC
AGCGAGAGCTCTTCTCTCACCGCTTCTAATCTTACCACCGTCGCCGCCGCTCTCCGTACC
GCGGCGGCGGAAGTTGCTTCGTCGGTTATTCCAGGTTTTGATATGCCTAATATGAAAATT

TACGGTAACGGGATCGAGTGGTTCGACGTTACTTGGACAAGGCTCATCGGCCGGTGGTGT
TTCTCGGAGATCGGTGGTTTCCGGCGGTTTCAGCTATTGAACTACACCGTTTGGATT
GGGGGTAAATTCGTAAATCAAGATGATCATCTGAAGTTAGAAGGTGAACTGTACAGCAG
CAACAGTTTGGAGATCGAACGGCTCAGGTTGAGTTTCAAGGAAGATCTTCGGATCCGAAT
ATGGGATTTGAACCGTTGGATTGGGGAAGTGGCGGTGGAGATCAAACACTGTTTGATT
ACCAGTACCGTTGATCATGCTACTGGAGTCAAAGTCAATGGACGTCGTCTGACCAAGAT
CAGAGTGGTCTCTACCTTCCTGA

>G1904 Amino Acid Sequence (domain in aa coordinates: 53-95)
MQDIHDFSMNGVGGGGGGGRFFGGGIGGGGGDRRMRAHQNNILNHHQSLKPCPNLSLN
TKFCYYNNYNLSQPRHFCKNCRRYWTGGVLRNVPVGGGCRKAKRSKTKQVPSSSSADKP
TTTQDDHHVEEKSSTGSHSSSESSSLTASNSTTVAAVSVTAAAEVASSVIPGFDMPNMKI
YNGIEWSTLLGQSSAGGVFSEIGGFPAVSAIETTPFGFGKFVNQDDHLKLEGETVQQ
QQFGDRTAQVEFQGRSSDPNMGFEPLDWGSGGDQTLFDLTSTVDHAYWSQSQWTSSDQD
QSGLYLP*

>G1906 (1..795)
ATGGTGGAACGTGCTCGGATCGCAAAAGTCCCATTGCCTGAAGCAGCTCTAAATTGCCCT
AGATGTGACTCAACCAATACTAAGTTCTGTACTTCAATAACTATAGCCTTACTCAACCT
CGCCATTTCTGCAAAACATGTCGTGCTATTGGACACGTGGCGGTTCTTGAGGAATGTT
CCTGTTGGAGGAGGCTTTAGGAGGAACAAGAGAAGCAAATCCAGATCGAAATCTACGGTC
GTGGTCTCGACTGATAATACTACTAGTACTTCATCACTTACTTCTCGCCCAAGTTACTCA
AACCTTAGCAAGTTTCATAGCTACGGTCAAATCCCGGAGTTTAATTCCAACTTGCCCATC
TTGCCTCCTCTCCAAAGCCTTGGAGATTACAATTCAAGCAACACTGGATTAGATTTTGGT
GGAACCTCAAATAAGCAACATGATAAGTGGTATGAGTTCTAGTGGTGGGATCTTGGATGCA
TGGAGAATACTCTCCATCACAAACAGCTCAGCAATTCCCTTTCTTGATCAACACTACCGGA
TTGGTGCAATCTTCAAACGCGTTATATCCATTACTAGAAGCGGGGTTAGCGCCACGCAA
ACAAGAAATGTGAAGGCGGAAGAGAATGATCAGGATCGGGGTAGGGATGGGGATGGAGTG
AATAACTTATCAAGAACTTTTGGGTAATATCAACATAAACTCAGGCAGGAACGAGGAA
TACACATCATGGGGAGGTAACAGTTCTTGGACCGGTTTCACCTCCAACAACTCAACAGGC
CATCTCTCATTCTAA

>G1906 Amino Acid Sequence (domain in AA coordinates: 19-47)
MVERARIAKVPLPEAALNCPDSTNTKFCYFNYSLTQPRHFCKTCRRYWTGGSRLNV
PVGGGFRNRKRSKRSKSTVVSTDNSTSSSLTSRPSYNSPKFHSYQIPEFNSNLPI
LPPLQSLGDYNSNTGLDFGGTQISNMISGMSSSGILDWRIPPSQQAQFPFLINTTG
LVQSSNALYPLLEGGVSATQTRNVKAEENDQDRGRDGDGVNLSRNLGNINININSRNEE
YTSWGGNSSWTGFTSNSTGHLSF*

>G1913 (1..744)
ATGGAGAGAGCAGGCGCTTGACATCATCGTTTATATGGCGGCCAAACGCAACGCAAC
GCGGAGATCAGCGGAGTTGTCCAAGATGTGGATCCTCTAACACAAAGTTCTGTACTAC
AACAACATAGCCTCACTCAGCCTCGCTACTTCTGCAAAGGCTGCCGCAGATATTGGACC
AAAGGTGGTTCCTCCGCAATGTTCTGTAGGCGGTGGCTGTGAAAATCCCGCCGCCCC
AAATCATCTTCTGGTAACAATACTAAACTAGCCTAACCGCTAATCTGGCAACCCCGGT
GGTGGTTACCAAGCATCGATCTTGCTCTTGTTTACGCCAATTTCTTGAATCCAAAGCCT
GACGAATCTATACTACAAGAAAATGCGACTTAGCCACTACGGATTTTTTGGTAGATAAT
CCTACCGGCACCTTCATGGACCTTCATGGAGTATGGACATCAATGATGGTCATCATGAT
CATTATATTAATCCGGTGGAACACATTGTGGAGGAATGTGGTTATAATGGCTTGCCCTCA
TTTCTGGTGAAGAGCTTCTCTTTAGACACTAATGGTGTGGTCTGATGCTTTGTTG
ATTGGTCATAACCATGTAGACGTTGGCGTGAATCCGGTTCAGGCTGTACACGAACCGGTG
GTTCAATTCGCTGAAGAATCCAATGATTCACCAATCTCTGTTTGAAGTTGGAGCCCT
TTTGATTTCACTGCCGATGGATGA

>G1913 Amino Acid Sequence (domain in AA coordinates: 27-55)
MERAELTSSFIWRPNANANAEITPSCPRCGSSNTKFCYYNNYSLTQPRYFCKGCRRYWT
KGGSLRNVVPVGGGCRKSRRPKSSGNNNTKSLTANSNPGGSGPSIDLALVYANFLNPKP
DESILQENCDLATDFLVDNPTGTSMDPSWSMDINDGHHDHYINPVEHIVEECGYNGLPP
FPGEELLSLDTNGVWSDALLIGHNHVDVGVTVPQAVHEPVVHFADESNDSTNLLFGSWSP
FDFTADG*

>G1914 (1..945)
ATGGAGAGATACAAGTGTAGATTTTGCTTCAAGAGCTTCATCAATGGAAGAGCTTTAGGT

GGTCACATGAGATCTCACATGCTTACTCTTTCTGCAGAACGTTGTGTAATAACTGGTGAA
GCAGAAGAAGAAGTAGAGGAACGGCCGAGTCAACTCTGTGACGACGACGATACCGAG
TCCGATGCTTCTTCTTCTTCTGGTGAGTTTGATAATCAAAGATGAATCGTCTTGATGAT
GAATTGGAGTTTGATTTCGCTGAAGACGACGACGTTGAAAGTGAAACCGAGTCGTCCAGG
ATTAACCCAACTCGGCGACGATCTAAACGAACCTCGGAAACTTGATCGTTTGATTTCGAC
TTTGAGAAGCTAACCAACGAGCCAACCCAGTGAGTTAGTGGCCGAGCCAGAGCATCACAGC
TCAGCTTCTGATACAACAACGGAGGAAGATCTCGCCTTTTGTCTCATTATGCTGTCCAGA
GACAAATGGAAGCAACAGAAGAAGAAGCAACGTGTAGAAGAAGATGAGACAGATCAT
GACAGTGAAGATTACAAATCAAGCAAGAGCAGAGGGAGATTCAAGTGTGAGACTTGTGGT
AAAGTGTTTAAATCTATCAAGCATTAGGAGGACACAGAGCAAGCCACAAGAAGAACAAG
GCATGCATGACGAAAACAGAGCAAGTTGAAACAGAGTACGTTCTTGAGTAAAGGAGAAG
AAAGTTCATGAATGTCCGATCTGTTTtagggTTTTACTTCAGGGCAAGCACTTGGAGGT
CATAAGAGATCTCAGGAAGTAACATCGGAGCAGGAAGAGGATTGTCTAGTAAGTCAAATT
GTCCAAATCGAAGAAGAAGTATCAGTGAAACAGAGGATGATTGATCTTAATCTTCCTGCA
CCTAATGAAGAAGATGAAACTTCTTTGGTGTGTGATGAATGGTGA
>G1914 Amino Acid Sequence (domain in AA coordinates:195-216, 245-266)
MERYKCRPFCFSFINGRALGGHMRSHMLTLAERCVITGEAEVEEVEERPSQLCDDDDDE
SDASSSSGEFDNQMNRLDDELEFDFAEDDDVESETESSRINPTRRRSKRTRKLSFDFD
FEKLTTSPSELVAEPEHHSSASDTTTEEDLAFCLIMLSRDKWKQKQKKKQKQVEEDET
DSEDYKSSKSRGRFCETCGKVFKSYQALGGHRAHSHKKNKACMTKTEQVETEVVLGVKEK
KVHECPICFRVFTSGQALGGHKRSHGSGNIGAGRLSVSVQIVQIEEVSVKQRMIDLNLPA
PNEEDETSLVFDEW*
>G1925 (1..945)
ATGGAGAAAATCTTCTCCGGGGTTCAGATTTTCATCCTACAGACGAGGAGCTCATAACG
CATTATCTATGTCCGAAAGTCTCCGATATAGGATTCACCGGTAAAGCTGTCTCGACGTT
GATCTCAACAAGTGTGAACCTTGGGATTTGCCAGCCAAGGCTTCAATGGGAGAGAAAGAG
TGGTATTTCTTCAGCCAAGGGATCGGAAATATCCAACCGGTTTAAAGACAAACCGGGCA
ACAGAAGCTGGTTACTGGAAAACACCGGGAAAGATAAAGAAATATACCGAAGTGGAGTG
TTGGTTGGGATGAAGAAAACCTAGTTTCTACAAAGGAAGAGCTCCCAAAGGTGAGAAA
AGCAATTGGGTTATGCTAGTACAGGCTTGAGAGCAAACAACCTTCAACCCACGAAT
AAGGAGGAATGGGTAGTGTGTAGGGTTTTCGAAAAGAGCACGGCAGCAAAGAAAGCACA
GAACAACAACCTCAATCTTCTCAACCATCTTTTGGATCTCCATGCGATGCAAACCTCATCA
ATGGCAAATGAGTTTGAAGATATTGATGAGCTTCCGAATCTGAATTCAAACTCATCAACC
ATCGATTACAATAATCATATCCATCAATATTTCGCAACGCAATGTTTACTCAGAAGACAAC
ACAACAAGTACGGCTGGTCTCAACATGAACATGAACATGGCTAGTACTAATCTTCAGTCT
TGGACAACAAGTCTCCTTGGTCCGCCTTTATCTCCAATCAACTCTTTGTTGCTCAAGGCT
TTCCAAATCAGGAACCTTATAGTTTCCCAAAGAGATGATCCCAAGTTTCAATCATTTCT
TCTCTTCAACAAGGAGTCTCCAATATGATCCAAAATGCTTCAAGTTCGTCTCAAGTGCAA
CCCCAACCGCAAGAGGAAGCGTTTAAATATGGACTCCATATGGTGA
>G1925 Amino Acid Sequence (conserved domain in AA coordinates:6-150)
MEENLPPGFRFHPTDEELITHYLCRKVSDIGFTGKAVVDVLDNKCEPWLPAKASMGKE
WYFFSQDRDKYPTGLRINRATEAGYWKTTGDKKEIYRSGVLVGMKKTLVFYKGRAPKGEK
SNWVMHEYRLSKQPFNPNTNKEEWVVCVFEKSTAAKKAQEQPQSSQPSFGSPCDANSS
MANEFEDIDELPNLNSNSTIDYNNHIIHQYSQRNVYSEDNTTSTAGLNMNMNMMASTNLQS
WTTSLGPPPLSPINSLLLKAFQIRNSYSFPKEMIPSFNHSSLQQGVSNMIQNASSSSQVQ
PQPQEEAFNMDSIW*
>G1929 (1..366)
ATGTGTAGAGGCTTGAATAATGAAGAGAGCAGAAGAAGTGACGGAGGAGGTTGCCGGAGT
CTCTGCACGAGACCGAGTGTTCGGTAAGGTGTGAGCTTTGCGACGGAGACGCCTCCGTG
TTCTGTGAAGCGGATCGCGCTTCTCTGTAGAAAATGTGACCGGTGGGTTTCATGGAGCG
AATTTTCTAGCTTGGAGACAGTAAGGCGCGTGTGCTATGCACTTCTGTGAGAACTCAGC
CGCCGGTGCTCGTCCGAGATCATGACTTCCACGTTGTTTTACCGTCCGTGACGACGGTC
GGAGAAACCACCGTGGAGAATAGAAGTGAACAAGATAATCATGAGGTTCCGTTTGTGTTTT
CTCTGA
>G1929 Amino Acid Sequence (domain in AA coordinates:31-53)
MCRGLNNEESRRSDGGGCRSLCTRPSVPVRCCLDGDASVFCEADSAFLCRKCDRWVHGA
NFLAWRHVRRVLCTSCQKLTRRCLVGDHDFHVVLPSVTTVGETTVENRSEQDNHEVPFVF

L*

>G1930 (76..1077)

ATTCACATTACTAATCTCTCAAGATTTTCAAAATTTTCTTGATTTTCTCTCAGTTTCTT
ATTTTCGTTTTCATAACATGGATGCCATGAGTAGCGTAGACGAGAGCTCTACAACACAGAT
TCCATTCCGGCGAGAAAGTCACTCGTCTCCGGCGAGTTTACTATATAGAATGGGAAGCGGA
ACAAGCGTGGTACTTGATTTCAGAGAACGGTGTGCAAGTCAAGTCAAGCCGAATCAAGA
AAGCTTCTTCTTCAAGATTCAAAGGTGTGTTCTCTCAACCAAATGGAAGATGGGGAGCT
CAGATTTACGAGAAACATCAACGCGTGTGGCTTGGTACTTTCAACGAGGAAGACGAAGCA
GCTCGTGCTTACGACGTCGCGGCTCACCGTTTCCGTGGCCGCGATGCCGTTACTAATTTT
AAAGACACGACGTTTCAAGAAGAGGTTGAGTTCTTAAACGCGCATTTCGAAATCAGAGATC
GTAGATATGTTGAGAAAAACACTTACAAGAAGAGTTAGACCAAAGGAAACGTAACCGT
GACGGTAACGGGAAAAGAGACGACGCGGCTTTGCTTTGGCTTCGATGGTGGTTATGACGGGG
TTTTAAACCGGCGAGTTACTGTTTGAAGAAACGGTAACGCCAAGTGACGTCGGGAAACTA
AACCCTTTAGTTTATACCAAAACACCAAGCGGAGAAACATTTCCGTTACCGTTAGGTAAT
AATAACGTCTCCGTTAAAGGTATGCTGTTGAATTTCAAGACGTTAACGGGAAAGTGTGG
AGGTTCCGTTACTCTTATTGGAATAGTAGTCAAAGTTATGTTGACCAAAGGTTGGAGT
AGATTCGTTAAAGAGAAGAGACTTTGTGCTGGTGGTATTTGATCAGTTTAAAGATCCAAC
GATCAAGATCAAAAATCTTTATCGGGTGGAAATCGAAATCCGGGTTGGATCTAGAGACG
GGTCGGGTTATGAGATTGTTTGGGGTTGATATTTCTTTAAACGCCGTCGTTGTAGTGAAG
GAAACAACGGAGGTGTTAATGTCGTCGTTAAGGTGTAAGAAGCAACGAGTTTGTATAAA
CAATTTAACACTTGGGAAAGAAAAAAGCTTTTGTATTTAATTTCTCTTCAACGTTA
ATCTTGCTGAGATTA

>G1930 Amino Acid Sequence (domain in AA coordinates: 59-124)

MDAMSSVDESSTTTDSIPARKSSSPASLLYRMGSGTSVVLDSENGVEVEVEAESRKLPS
RFKGVVPPQNGRWGAQIYEKHQVRVWLGTFFNEEDEAARAYDVAHRFRGRDAVTNFKDTTF
EEVEFLNAHSKSEIVDMLRKHTYKEELDQKRNRDNGKETTAALASMVVMTGFKTAE
LLFEKTVTPSDVGVKLNRLVIPKHQAQKHPLPLGNNNVSVKGMMLNFEDVNGKVWRFRYS
YWNSSQSYVLTKGWSRFVKEKRLCAGDLISFKRSNDQDQKFFIGWKS KSGLDLETGRVMR
LFGVDISLNAVVVVKETTEVLMSLRCKKQVRV*

>G195 (51..1031)

TTTTCTTTTCTTTCTTTTGGTTTAAAGTTTTTCTCTTTGTTCTTCGTCATGTCTCATG
AAATCAAAGATCTTAACAACTACTACTACATTCATCGTATAATCATTACAATATCAACA
ACCAAAATATGATTAATCTCCCTTACGTTTCTGGTCCATCTGCTTATAATGCAACATGA
TCTCATCATCACAAGTAGGTTTGTATCTACCTCGAAGAACTTGAGTCCTCAAGGAGCCT
TCGAGTTGGGTTTCGAGCTTCTCCATCTTCTTGACTTTTTTAATCCTTCCCTCGATC
AAGAGAACGGTTGTATAATGCTTATAATTATAATAGTAGTCAAAAGAGTCATGAAGTTG
TCGGTGATGTTGTGCAACCATTAAAGAGTGAAGTTAGGGTTTCAGCATCTCCTTCTTCAA
GTGAGGCCGATCATCATCCAGGAGAAGATTCCGGCAAGATCCGGAAGAAAAGAGAAGTTC
GCGATGGAGGAGAAGATGATCAACGCTCTCAGAAAGTAGTTAAACAAAGAAGAAAGAGG
AGAAGAAAAAAGAGCCACGAGTCTCGTTTCATGACTAAGACCGAAGTTGATCATCTCGAAG
ACGGCTATCGTTGGAGAAAGTATGGCCAAAAAGCAGTCAAAAACAGTCCTTATCCGAGGA
GTTACTATAGATGCACGACTCAGAAGTGCAACGTGAAGAAGAGAGTGGAGAGATCTTACC
AAGACCAACGGTCTGTCATCACAACCTACGAGAGTCAACACAACCATCCGATCCCGACCA
ATCGTCGGACAGCAATGTTCTCTGGAACCCCGCATCTGATTATAACCCATCATCGTCTC
CAATATTCTCCGATCTCATCATCAATACTCCAAGAAGCTTCTCAAATGATGATCTCTTCC
GTGTGCCATACGCTAGTGTGAACGTGAACCTAGTTATCATCAACAGCAACATGGATTTC
ATCAACAGGAGAGTGAGTTTCGAGCTCTTGAAGGAGATGTTTCCTTCGGTTTTCTTCAAAC
AAGAGCCTTGATGATATAATATAATATAGAAACAATTTTTTTCTGCTAAGAAATATAGA
ACAAAACCTGGATGCATAATAAGTGATGATAGTGTTATTTATTTTTTGCATGTATATATT
ATACATGTTTTGTTAACTAGCTATAGGATATACTGGTAGTAATTAAGCATAAATATGGAG
CCCTTCGACTTATTACAATAATTTTGGTATGGAAAAANTNGNTACATGCCTGCCTTTT
NNNTTNNNG

>G195 Amino Acid Sequence (domain in AA coordinates: 183-239)

MSHEIKDLNNYHYTSSYNHYNNQNMNINLPYVSGPSAYNANMISSSQVGFDLPSKNLSP
QGAFELGFELSPSSSDFFNPSLDQENGLYNAYNINSSQKSHEVVGDGCATIKSEVRVSAS
PSSSEADHHPGEDSGKIRKKREVRDGGEDDQRSQKVVKTKKKEKKKEPRVSFMTKTEVD
HLEDGYRWRKYGQKAVKNSPYPRSYRCTTQKCNVKKRVERSYQDPTVVITTYESQHNP

IPTNRRRTAMFSGTTASDYNPSSSPIFSDLIINTPRSFSDNDLFRVPYASVNVNPSYHQQQ
HGFHQESEFELLKEMFPSVFFKQEP*

>G1954 (196..1440)

ATTTATGACTTCTCAATACAAAAAGCTCCCTCACTTTTTTAAGTTTGTCTTCTCTAAT
CCGTCTTCTTCTACTATCTTGCATGTCTTGGCTCTTTATATACATCTCTCGTAAACCCT
AGCAAATCATACAAGGTCAAGAAGCTTGACCTTCATTAGACTTAAGCAGTTTATAATCAA
CTACCACGAATAGCAATGGATAAAGATTACTCGGCACCAAACTTCTTAGGTGAATCCTCA
GGCGGTAAACGATGATAACAGCTCTGGTATGATAGACTATATGTTCAATAGAAACCTTCAA
CAACAACAAAAGCAATCGATGCCACAACAGCAGCAACATCAACTCTCTCTTCCGGATTT
GGAGCAACACCCTTTTGATAAAATGAAGCTTCTCTGATGTGATGCAGTTTGCGGACTTCGGT
TCGAAACTTGCGTTGAACAGACCAGAAACCAAGACGATCAAGAAACCGGGATTGACCC
GTTTATTTCTTGAAGTTCCTGTCTTGAACGACAAAATAGAGGACCATAACCAAACCCAA
CATCTCATGCCTTCTCATCAGACGTCTCAAGAAGGAGGTGAGTGTGGAGGAAACATAGGC
AATGTGTTTCTTGAAGAAAAAGAAGATCAAGACGATGACAACGACAACAACCTCCGTGCAA
CTACGTTTTATTGGAGGAGAAGAAGAAGATAGGGAGAACAAGAATGTTACGAAAAAGGAG
GTGAAGAGCAAGAGGAGAGAGCTAGAACGAGCAAGACCAGCGAAGAAGTGGAAAGCCAA
CGGATGACTCATATCGCGTTCGAAAGAAACCGTAGGAAGCAATGAATGAGCATCTTCGT
GTCCTTAGATCTCTCATGCCTGGCTCCTACGTTCAAAGGGGAGACCAAGCGTCAATCATA
GGAGGAGCAATAGAGTTTGTGAGAGAGCTCGAGCAACTCCTACAATGTCTTGAATCACAG
AAGCGTCGAAGAATCTTAGGAGAAACCGTAGGGACATGACAACGACAACGACTTCTTCT
TCTTCTCCCATAACTACGGTAGCGAACCAAGCACAACCGCTCATTATTACGGGAAATGTA
ACCGAGCTAGAGGGCGGAGGAGGGCTTCGGGAGGAGACTGCGGAGAACAAAGTCGTGCTTG
GCTGACGTGGAGGTGAAGCTGCTAGGGTTTGACGCCATGATCAAGATACTTTCAAGAAGA
AGGCCGGGACAGCTGATTAAAGACTATAGCTGCTTTGGAGGATCTTCATCTCTCTATTCTT
CACACTAACATCACTACCATGGAACAAACCGTCTCTACTCCTTTAATGTCAAGATAACA
AGTGAACGAGGTTTACGGCAGAAGACATAGCAAGTTCATCCAACAGATATTTAGTTTC
ATTTCATGCAAAATACCAACATATCTGGAAGCTCTAACCTGGGAAATATTGTGTTTACTTGA
AAATCATCACACGGCGACAACCTTTGTACACTGGTGAAGATTACAGTACGTAATAATCTCT
ACATATTGGGTTTTATTCTCCAAGCATTTGGAAGAGTGTAAAGTTAAAGGGAGTGCTTA
CTTTATTTTTTTGGGGCTTTTTTCATGCAATTTAAATTTTAGTGTGATTGTGTGCTTG
TAATGTTAGAATCTGTTGTTGTGATTCTGCTGCTTTGATTGTAGGTTTTGAACAAGCG
GTTTAGAATGCTAAACCCTTATTACTTGAAATAACTTTTTTCACAAAAA

>G1954 Amino Acid Sequence (domain in AA coordinates:187-259)

MDKDYSAPNPLGESSGNDNDSGMDIDYMFNRNLQQQQKQSMPPQQQHQHLSPSGFGATPF
DKMNFSDVMQFADFSGKLALNQTRNQDDQETGIDPVYFLKFPVLNDKIEDHNQTHLMPS
HQTSQEGGECGNIGNVFLKEEDQDDNNDNSVQLRFIGEEEDRENKNVTKEVKSKR
KRARTSKTSEEVESQRMTHIAVERNRRKQMNEHLRVLRLSLMPGSYVQRGDQASIIGGAIE
FVRELEQLQLCLESQKRRRILGETGRDMTTTTSSSPITTVANQAQPLIITGNVTELEG
GGGLREETAENKSLADVEVKLLGFDAMIKILSRRRPGQLIKTIAALEDLHLSILHTNIT
TMEQTVLYSFNVKITSETRFTAEDIASSIQQIFSFHANTNISGSSNLGNIVFT*

>G1958 (107..1336)

GTACCGTCGACCGATTATCCCCAAGAGGAGAATCCTCATAATCATTTTCTCCGATTCGAT
TCGTCTTCCCTGGTCCCTGGATTGCTTCATGAATTTCTAGGACAACAATGGAGGCTCGTCC
AGTTCATAGATCAGGTTTCGAGAGACCTCACACGCACTTCTTCAATCCCATCTACACAAAA
ACCTTCACCAGTAGAAGATAGTTTCATGAGATCAGATAACAACAGTCAGTTAATGTCTAG
ACCATTAGGACAAACCTACCATTTACTTTCATCTAGTAACGGTGGAGCTGTTGGACATAT
ATGTTCTTCTTCATCATCTGGTTTTGCAACCAATCTCCATTACTCAACTATGGTATCTCA
TGAGAAACAACAACACTACACAGGAAGCAGCAGTAATAATGCTGTGTCAGACACCAAGCAA
CAACGATAGTGCTTGGTGTGATGATTCATTGCCAGGAGGGTTTCTTGACTTCCATGAAAC
CAACCCGGCGATTCAAAAACACTGTGATGAGGATGGTGGCATTGCGGCTGCTTTGA
TGACATTCAAAAACGAAGTATTGGCATGAATGGGCTGACCATTGATCACTGATGATGA
TCCTTTGATGTCTACTAACTGGAATGATCTCTTGCTTGAAACAAATCCAATTGATTC
AAAGGACCAGAAGACTGCAAATTCGCAACCTCAGATTGTTTCAGCAGCAACCTTCTCC
GTCTGTGGAATTGCGACCTGTTAGCACACATCTTCAAACAGCAATAACGGAACGGGCAA
GGCAGCAATGCGTTGGACGCCAGAGCTTACGAGGCTTTGTTGAGGCTGTCAACAGTCT
TGGCGGTAGTGAAAGAGCTACTCTAAAGGGGTACTGAAGATTATGAAAGTTGAAGGCTT

GACTATATATCATGTAAAGCCATTTACAGAAATATAGGACAGCTAGATATCGGCCAGA
ACCATCAGAACTGGTTCGCCAGAAAGGAAGTTGACACCGCTTGAACATATAACATCTCT
TGATTTGAAAGGTGGGATAGGTATTACAGAGGCTCTACGACTTCAGATGGAAGTACAGAA
GCAACTCCATGAGCAGCTCGAGATTCAAAGAAACCTGCAACTCCGAATAGAAGAACAAAG
CAAGTACCTGCAATGATGTTTCGAGAAGCAAACTCTGGTCTTACCAAAGGGACAGCCTC
AACATCAGATTCCGAGCCAAATCTGAACAAGAAGACAAGAAGACTGCTGATTGGAAGGA
GGTTCAGAGAAGAAACCAGGAAATGTGAGGAAGTAGAATCTCCACAGCCAAAGCGTCC
CAAAATCGATAATTGAAAGTATTGGTCTTTTGCTGATAATCTCGAGTTTCAGAGTTAA
CAGTGATAGAGAGAACGAGCTCTTATCTTGAGGTTCTTCAGGACTTCTCTCGCGGCCGCT
CTAG

>G1958 Amino Acid Sequence (domain in AA coordinates: 230-278)

MEARPVHRSGSRDLTRTSSIPSTQKPSPVEDSFMRSDMNSQLMSRPLGQTYHLLSSSNGG
AVGHICSSSSSGFATNLHYSTMVSHEKQQHYTGSSSNNAVQTPSNNDASWCHDSLPGGFL
DFHETNPAIQNNCQIEDGGIAAFDDIQKRSDWHEWADHLITDDPLMSTNWNLDLLETN
SNSDSKDQKTLQIPQPQIVQQQSPSVELRPVSTSSNSNNGTGKARMRWTPELHEAFVE
AVNSLGGSERATPKGVLMKVEGLTIYHVKSHLQKYRTARYRPEPSETGSPERKLTPL
HITSLDLKGGIGITEALRLQMEVQKQLHEQLEIQRLQLRIEEQGYLQMMFEKQNSGLT
KGTASTSDSAKSEQEDKKTADSKEVPEETRKCCELESPPKRPKIDN*

>G196 (111..1421)

TCGACATCAGATTTCTCTCAGGATTCTTAATCATTTTTATTATATTGGATATTTGCTA
ATTCTTCCCGTGATAAATCTCATATAAACACGCATCATACATATATATTATGTGCAGCG
TCTTTGAGTTTCAAGACATGGACAACCTCCAAGGAGATCTAACAGACGTCGTACGAGGAA
TAGGATCAGGCCACGTTCCACCATCTCCTGGACCACCGGAAGGTCCATCTCCGAGCAGCA
TGTCTCCGCCGCCAACATCAGATCTCCACGTGGAATTCCTCCGCCGCTACTCTGCCA
GCTGTCTCGCAAATCCCTTCGGAGACCCGTTTCGTAAGCATGAAGGATCCTCTCATCCACC
TCCCGGCCAGCTACATCTCCGGCGCCGGTGATAATAAAGCAACAAAAGTTTTGCAATCT
TTCCAAAGATTTTTGAGGATGATCATATTAAGAGTCAATGCAGTGTCTTCCCAAGAATTA
AGATCTCGCAAAGTAACAATATCCACGATGCCTCCACGTGTAATTCTCCGCCATAACCG
TCTCCTCTGCCGCCGTAGCAGCTTCGCCGTGGGGCATGATCAACGTTAATACCACTAACA
GTCCAAGAACTGTTTACTTGTGCGATAATAATAACAACACGTATCATGCTCACAGGTTTC
AGATCTCTTCTTCCCTCGGAATCTCGGAATTAAGAGAAGGAAGAGCCAGGCAAAGAAAG
TGGTGTGCATACCGGCTCCAGCGCTATGAACAGCCGGTCCAGTGGAGAAGTTGTTCCGT
CTGATCTATGGGCTTGGCGAAAGTACGGTCAAAAACCTATCAAAGGTTCTCCTTATCCAA
GGGGTTACTACAGATGTAGCAGCTCAAAGGTTGTTTCAGCTAGGAAACAAGTCGAACGTA
GCCGCACTGATCCAAACATGTTAGTCATTACTTACACCTCTGAGCATAACCACCCATGGC
CTACTCAACGCAACGCTCTCGCAGGTTCCACTCGTTCCTCTTCTCCTCCTCTTTAAACC
CTTCTTCCAAATCCTCAACCGCAGCCGCCACTACTTCTCCCTCATCCAGATTTTTCCAAA
ACAACAGCAGCAAAGACGAACCCAATACTCCAATTGCTTCTCTTCCACTCATCCTC
CTTTTGACGCCGCCGCAATTAAGGAGGAGAACGTGGAAGAGCGTCAGGAAAAGATGGAGT
TCGATTATAATGACGTTGAAAATACCTATAGACCGGAGTTGTTGCAAGAGTTTCAACATC
AGCCGGAGGATTTCTTTGCCGATCTCGACGAGCTTGAGGGAGATTCTTTGACTATGTTGC
TCTCTCACAGTAGCGGCCGAGGCAACATGGAAAACAAAACGACGATTCCAGACGTTTTTA
GTGATTTCTTTGACGACGACGAGTCCTCAAGGTGCTTATAAATATTGTTGTTAATGTATA
CATAGAAATGAAATTATTCATGTAATTCGTTTTGTGTTAAATGACGGTATTTGCCTTTGC
A

>G196 Amino Acid Sequence (conserved domain in AA coordinates: 223-283)

MCSVFEFQMDNMFQGDLDVVRGIGSGHVSPPSGPPEGSPSSMSPPPTSDDLHVEFPSAA
TSASCLANPFGDPFVSMKDPLIHLPASYSIGAGDNKSNKSFAIFPKIFEDDHIKSQCSVF
PRIKISQSNNIHDASTCNSPAITVSSAAVAASPWGMINVNTTNSPRNCLLDVNNNTSSC
SQVQISSSPRLGIKRKSQAKKVVCIPAPAAMNSRSSGEVVPDDLWAWRKYQKPIKGS
PYPRGYRCSKSGCSARKQVERSRTDPNMLVITYTSEHNHPWPTQRNALAGSTRSSSSS
SLNPSSKSSATAATTSPSSRVFQNNSSKDEPNNSNLPSSTHPPFDAAAIKEENVEERQE
KMEFDYNDVENTYRPELLQEFQHQPEDFFADLDELEGLSLTMLLSHSSGGGNMENKTTIP
DVFSDFDDDESSRSL*

>G1965 (1..609)

ATGGATAACTTCAATGTTGTTGCCAATGAAGACAATCAAGTGAATGATGTGAAGCCTCCA
CCACCCCCACCGCAGTGTGTGCAAGATGTGATTCTGATAACACAAAATTTGTTACTAC

AACAATTATAGTGAGTTTCAACCGCGCTACTTCTGCAAGAACTGTGCAAGATACTGGACT
CATGGTGGGGCTTTAAGAAACGTACCAATTGGTGGGAGTAGTCGTGCCAAGCGGACAAGG
ATAAATCAACCTTCAGTTGCTCAGATGGTTCTGTTGGAATCCAACCAGGGAACCGTTT
AGTTCTTTGTCTCATATTCATGGTGGTATGGTAACAAATGTGCATCCAACCTCAAACCTTT
CGACCAAATCATCGCCTAGCTTTCCATAATGGATCATTTGAGCAAGATTATTATGATGTT
GGGTCTGATAATCTTTTGGTAAACCAACAAGTTGGTGGATATGTTGATAATCACAACGGT
TATCACATGAATCAAGTGGATCAATACAACCTGGAACCAGAGCTTCAATAACGCTATGAAC
ATGAATTATAATAACGCTAGCACTAGCGGAAGGATGCATCCTAGTCATTTAGAGAAGGGT
GGTCCTTGA

>G1965 Amino Acid Sequence (domain in AA coordinates:27-55)
MDNFNVVANEDNQVNDVKPPPPPRVCARCDSDNTKFCYNNYSEFQPRYFCNKRCRRYWT
HGGALRNVPIGSSRAKRTRINQPSVAQMVSVGIQPGNRFSSLSHIHGGMVTNVHPTQTF
RPNHRLAFHNGSFEQDYDVGSDNLLVNQQVGGYVDNHNHGYHMNQVDQYNWNQSFNNAMN
MNYNNASTSGRMHPSHLEKGGP*

>G1976 (1..1152)
ATGACTGATCCTTATTCCAATTTCTTCACAGACTGGTTCAAGTCTAATCCTTTTCACCAT
TACCCTAATTCCTCCACTAACCCCTCTCCTCATCCTCTCCTCCTGTACTCCTCCCTCT
TCCTTCTTCTTCTTCCCTCAATCCGGAGACCTCCGCGCTCCACCGCGCCACCAACTCCT
CCTCCTTCTCCTCCTCTCCGAGAAGCCCTCCCTCTCCTCAGCCTCAGCCCCGCCAACAAA
CAACAAGACCACCATCACACCATGACCACCTTATTCAAGAACCACCTTCAACCTCCATG
GATGTCGACTACGATCATCACCATCAAGATGATCATATAACCTCGATGACGATGACCAT
GACGTCACCGTTGCTCTTCACATAGGCCCTTCAAGCCCTAGTGCTCAAGAGATGGCCTCT
TTGCTCATGATGCTTCTTCTTCTCTCTCCTCGAGGACCACTCATCATCAGAGGACATG
AATCACAAGAAAGACCTCGACCATGAGTACAGCCACGGAGCTGTGCGAGGAGGAGAAGAT
GACGATGAAGATTGAGTCGGCGGAGACGGCGGCTGTAGAATCAGCAGACTCAACAAGGGT
CAATATTGGATCCCTACACCTTCTCAGATTCTCATTGGCCCTACTCAGTTCTCATGTCCT
GTTTGCTTCAAAACCTTCAACAGATACAATAACATGCAGATGCATATGTGGGGACATGGA
TCACAATACAGAAAAGGACCTGAATCTCTAAGGGGAACACAACCAACGGAATGCTAAGG
CTTCCGTGCTATTGCTGCGCCCCAGGCTGTGCGCAACAACATTGACCATCCAAGGGCAAAG
CCTCTCAAAGACTTTCAGAACCTTCAAACACATTACAAGAGAAAACATGGGATCAAACCT
TTCATGTGTAGAAATGTGAAAAGGCTTTCGCACTCCGAGGGGACTGGAGAACACATGAG
AAGAATTGTGGCAAACCTTTGGTATTGCATATGTGGATCTGATTTCAAGCACAAGAGATCT
CTCAAAGATCACATCAAGGCTTTTGGGAATGGTCATGGAGCCTACGGAATTGATGGGTTT
GATGAAGAAGATGAGCCTGCCTCTGAGGTAGAACAATTAGACAATGATCATGAGTCAATG
CAGTCTAAATAG

>G1976 Amino Acid Sequence (domain in AA coordinates: 219-323)
MTDPYSNFFTDFWFKSNPFHHYPNSSTNPSPHPLPPVTPPSSFFFFPQSGDLRRPPPPPTP
PPSPPLREALPLLSLSPANKQQDHHNHDHLIQEPPSTMDVDYDHHHQDDHNLDDDDH
DVTVALHIGLPSPSAQEMASLLMSSSSSSSRTHHHEDMNHKKDLDEYSHGAVGGGED
DDEDSVGGDGGCRISRLNKGQYWIPTPSQILIGPTQFSCPVCFKTFNRYNNMQMHMWGHG
SQYRKGPESLRGTQPTGMLRLPCYCCAPGCRNNIDHPRAKPLKDFRTLQTHYKRKHGIKP
FMCRCCKGKAFVRGDRWTHEKNCGKLWYICGSDFKHKRSLKDHIKAFNGHGHAYGIDGF
DEEDEPASEVEQLDNDHESMQSK*

>G2057 (27..1289)
GCCGTCTCGACGAATATGCTTACCAATGTCTGACGACCAATTCCATCACCCGCGCCTC
CTTCTTCAATGAGGCACCGTTCTACGTCGGATGCGGCGGACGGCGGCTGCGGCGAGATTG
TTGAGGTGCAAGGTGGTCACATTGTTTCGGTCTACCGGAAGAAAAGACCGCCACAGCAAAG
TCTGCACGGCTAAAGGGCCACGTGACCGGCGCGTGAGACTCTCTGCTCACACGGCGATT
AGTTTTACGATGTTCAAGACAGGCTTGGTTTCGACCGACCTAGCAAAGCCGTTGATTGGC
TTATCAAAAAGGCTAAGACTTCCATTGACGAGCTCGCTGAGCTTCTCCTGGAATCCCG
CCGATGCAATTCCGCTAGCCGCTGCTAACGCTAAACCCAGAAGAACCACCGCCAAAACCC
AAATCTCTCCGTCTCCGCCACCGCGCAACAGCAACAACAACAACAGCTTCAGTTTCG
GTGTTGGCTTCAACGGAGGAGGAGCAGAGCATCCGAGTAACAACGAGTCGAGTTTCTCC
CGCCGTCAATGATTGAGATTCGATAGCTGACACTATAAAGTCGTTTTTCCGGTGATTG
GCTCTTCAACGAGGCTCCTTCGAATCATAACCTTATGCACAACATCATCATCAGCATC
CGCCGGATTGCTTTCTCGAATAATAGCCAAAACCAAGATCTCCGTCTCTCGCTGCAAT
CGTTCCCGGATGGTCCACCGTCGCTTCTGCACCACCAACATCACCACCACACCTCTGCTT

CCGCCTCCGAGCCTACTCTGTTCTACGGACAGAGCAATCCGTTAGGGTTTGACACATCGA
GTTGGGAGCAGCAGTCGTCGGAATTCGGAAGGATTAGAGACTAGTGGCTTGAACAGCG
GCGGTGGCGGCGGAGCAACCGATACAGGAAACGGAGGAGGGTTCTGTTGCTCCTCCTA
CTCCTTCAACGACGTCGTTTCAGCCAGTTCTTGCCAAAGCCAACAGCTTTATTTCTCAGA
GGGGTCCCCTTCAGTCCAGTTCAGTCCCAGTCCGTCGTTGTTGATCCTCACCATC
ATCACCATTCATCTCCACCGACGATCTCAACCACCACCATCACCTTCTCCACCGGTTT
ACCAATCAGCAATCCCCGAATCGGATTCGCCTCAGGTGAATTCTTTCGGGTTTTCGCA
TACCAGCAGGTTTTCAGGGCCAAGAAGAGGAGCAGCAGCAGGTTCTCACTACAAGCCGT
CCTCTGCTTCTCTATTCTCGCCATTGACAATCGAACTAATCCTC

>G2057 Amino Acid Sequence (domain in AA coordinates: TBD)

MSDDQFHHPPPPSSMRHRSTSDAADGGCGEIVEVQGGHIVRSTGRKDRHSKVC TAKGPRD
RRVRLSAHTAIQFYDVQDRLGFDRPSKAVDWLIKAKTSIDELAE LPPWNPADAIRLAAA
NAKPRRTTAKTQISPSPPPPQQQQQQQLQFGVGFNGGGAEHPSNNESSFLPPSMDSDSI
ADTIKSFPPVIGSSTEAPSNHNLHNYHHQHPPDLLSRTNSQNQDLRLSLQSPDGP PSL
LHHQHHTSASASEPTLFYQSNPLGFDTSWEQSSSEFGRIQRLVAWNSGGGGGATDT
GNGGGFLFAPPTPSTTSFQVLGQSQQLYSQRGPLQSSYSPIRAWFDPHHHQSISTDD
LNHHHHLPPP VHQSAIPGIGFASGEFSSGFRIPARFQGEQEHDLTHKPSSASSISRH

*

>G2107 (79..624)

ACCACAAAACAGAGCAACACACAACACAAAGCTTCATTTC AATTCTGTTTCGAGAACCCT
TTGAGAACCAGATCGGAGATGGAACGACGATATCACCGTGGCGGAGATGAAGCCAAAG
AAGCGTGCTGGACGAGGATTTCAAGGAGACACGTCACCAATCTACAGAGCGTGCGG
CGTAGGGACGGCGACAAATGGGTATGCGAAGTCCGTGAACCGATTATCAGCGTCGAGTC
TGGCTCGGAACTTATCCGACGGCAGATATGGCCGACGTCACGACGTGGCGGTTCTT
GCTCTGCGCGGAGATCCGCGTGTGTAATTTCTCCGATTCTGCTTGGAGGTTGCCGGTG
CCGGCATCCACTGATCCGGACACGATCAGGCGCACGGCGCCGAAGCAGCGGAGATGTTT
AGGCCGCGGAGTTTAGTACAGGAATTACGGTTTACCCTCAGCCAGTGAGTTTGACACG
TCGGATGAAGGAGTCGCTGGAATGATGATGAGGCTCGCGGAGGAGCCGTTGATGTCGCCG
CCAAGATCGTACATTGATATGAATACGAGTGTGTACGTGGACGAAGAAATGTGTTACGAA
GATTTGTCACTTTGAGATTACTAAAATACGTATGTGTTAAAAAACCAAGATCGTATGTG
TATGTATGCATAATAAATGGGCTTAATGATGGGCATAGATATGATAGGTCCAGCCTATAT
GTTAAATGTGTTTTATTTTTGGTTTATCTAGTTTCTAGGTATTTACCAATTGTATTA
GTATAAGTTTATTATAAGAAATAATCAAAAATGTTGTTGCCAAAAA AAAAAAAAAA
AAAAA

>G2107 Amino Acid Sequence (domain in AA coordinates: TBD)

MENDDITVAEMKPKKRAGRRIFKETRHPYIRGVRRRDGDKWVCEVREPIHQRRVWLGTYP,
TADMAARAHDVAVLALRGRSACLNFSDSAWRLPVPASTDPDTIRRTAAEAAEMFRPPEFS
TGITVLP SASEFDTSDGAVAGMMMLAEPELMSPPRSYIDMNTSVYVDEEMCYEDLSLWS
Y*

>G211 (1..750)

ATGATGTCATGTGGTGGGAAGAAGCCAGTGCTTAAGAAAACAACGCCGTGTTGCACGAAG
ATGGGGATGAAGAGAGGACCATGGACGGTGGAGGAAGACGAGATTCTTGTGAGCTTCATT
AAGAAAGAAGGTGAAGGACGGTGGCGATCGCTTCTTAAGAGAGCTGGTTTACTCAGATGT
GGAAAGAGCTGTGCTCTACGGTGGATGAACTATCTCCGACCCTCGGTTAAACGTGGAGGA
ATTACGTCCGACGAGGAAGATCTCATCTCCGTCTTACCGCCTCCTCGGCAACAGGTGG
TCATTGATCGCGGGAAGGATACCGGGAAGGACTGATAATGAAATTAAGAATATTTGGAAC
ACTCATCTTCGTAAGAACTTTTAAGGCAAGGAATTGATCCTCAAACCCACAAGCCTCTT
GATGCAACCAACATATAAACCAGAAGAAGAAGTTCCGGTGGACAAAAGTACCCTCTA
GAGCCTATTTCTAGTTCTCATACTGATGATACCACTGTTAATGGCGGGGATGGAGATAGC
AAGAACAGTATCAATGTCTTTGGTGGTGAACACGGCTACGAAGACTTTGGTTTCTGCTAC
GACGACAAGTTCTCATCGTTTCTTAATTCGCTCATCAACGATGTTGGTGATCCTTTTGGT
AATATTATCCCAATATCTCAACCTTTGCAGATGGATGATTGTAAGGATGGGATTGTTGGA
GCGTCGTCTTCTAGCTTAGGACATGACTAG

>G211 Amino Acid Sequence (conserved domain in AA coordinates: 24-137)

MMSCGGKKPVSKTTPCCTKMGMRGPWTVEEDELVSFIKKEGEGRWRS LPKRAGLLRC
GKSCRLRWNNYLRPSVKRGGITSDEEDLILRLHRLGNRWSLIAGRIPGRTDNEIKNYWN
THLRKKLLRQGIDPQTHKPLDANNIHKPEEEVSGGQKYLEPISSHTDDTTVNGGDGDS

KNSINVFGEHGYEDFGFCYDDKFSSFLNSLINDVGD PFGNIIPISQPLQMDDCKDGIIVG
ASSSSLGHD*

>G2133 (26..457)

ATCTCATCTTCATCCACCCAAAAACATGGATTCAAGAGACACCGAGAACTGACCAGAG
CAAGTACAAAGGTATCCGTCGTCGGAATGGGGAATGGGTATCAGAGATTCGTGTCCC
GGGAACCTCGTCAACGTCTCTGGTTAGGCTCTTTCTCCACCGCAGAAGGCGTGCCGTAGC
CCACGACGTCGCTTTTACTGCTTGACCGACCATCTTCCCTCGACGACGAATCTTTTAA
CTTCCCTCACTTACTTACAACCTCCCTCGCCTCCAATATATCTCCTAAGTCCATCCAAAA
AGCTGCTTCCGACGCCGGCATGGCCGTGGACGCCGGATTCCATGGTGCTGTGTCTGGGAG
TGGTGGTTGTGAAGAGAGATCTTCCATGGCGAATATGGAGGAGGAGACAACTTAGTAT
CTCCGTGTATGATTATCTTGAAGACGATCTCGTTTGATCTATACGAGTACGTTTGTAGCA
GTAA

>G2133 Amino Acid Sequence (domain in AA coordinates:11-83)

MDSRDTGETDQSKYKIGIRRRKWKVWSEIRVPGTRQRLWLGSFSTAEGAAVAHDVAFYCL
HRPSSLDDESNFNPHELLTSLASNISPKSIQKAASDAGMAVDAGFHGAVSGSGGCEERS
MANMEEEDKLSISVYDYLEDDL*

>G2134 (36..644)

GAGCAAAAACTTTGTGTGCGTGTGTGTGTGTGTTCATGGCTGGTCTTAGGAATTCGGTA
ACAGCGACAAAGCGCAAAACGATGGCAAAGGTGTACCATCTGCCTACAGAGGAGTCCGGA
AGAGAAAATGGGGGAAATGGGTGTCTGAAATCCGTGAACCGGGGACCAAGAACCGTATCT
GGCTAGGCAGTTTCGAGACTCTGAAATGGCTGCAACCGCATACGACGTGGCAGCATTTC
ATTTTCAGAGGGAGAGAAGCTCGTCTCAACTTCCCTGAGCTCGCCAGCAGCCTTCCACGTC
CTGCAGACTCTAGCTCAGACAGCATTCGCATGGCAGTTTCATGAGGCAACACTCTGCCGCA
CCACCGAAGGAACAGAGTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA
CAACAATGGTCAGACTCTCGCCAGGGAAATTCAAGCGATCAACGAGTCAACTTTGGGAT
CTCTACTACAATGATGCATTCAACATACGACCTATGGAGTTTGCTAATGATGTGGAGA
TGAATGCTTGGGAAACATACAGAGTGACTTCTTTGGGACCTTAACCCCAAACCTAA
CTCATGGAGAGCTTCTACAGCTCAATCTTACAATACAGCATAAGTTACTGGCTTAGAAT
ACTTAAATTTATTGAAGTTTAGTTTTCAGAGTCTACCACAAGGGTTGTTGATTCTGACGT
TATAGCAAAGAATAAAGCTCATCAGATTTTGGAGGGAAGACTCTATGAGCTTGATGGGT
CCCTGAAAGGACCTCTTACAAATATTTTAAATTTTTTTGTTACTAGTAGAAACATAGA
TTATGAGGTGTGACTTATTATTATTTTTTACAATTGTTTGTACCTCATTGATGTATTTG
ATT

>G2134 Amino Acid Sequence (domain in AA coordinates: TBD)

MAGLRNSGNSDKAQNDGKGVPSAYRGVVRKRKWKVWSEIREPGTKNRIWLGSFETPEMAA
TAYDVAAFHFRGREARLNFPELASSLPRPADSSSDSIRMAVHEATLCRTTEGTESAMQVD
SSSSSNVAPTVMRLSPREIQAINESTLGSPTMMHSTYDPMFANDVEMNAWETYQSDFL
WDP*PQNLTHGELLQLNLTI*

>G2151 (236..1321)

TTTTTTTTTTAGGGTTCATAAGAACAAATTGGATTTTGAGCTCACAGTATAAATAACCCG
ACTTTGATTACTGGGTAAATTTAAACCGCCATTGTTGTCTCTTTACTACTTTTGGGAA
TTAGGGTTTATGATTTCTGGGTATTAGATTAGATAAATTGTTTTCTTTTGTGTTAATC
AATTTAAAAATCTCTTATTTCTGTTAAAGACTTGTAATTTTGGAGTTTTAATGCATGGA
CGGAAGAGAAGCAATGGCATTTCCAGGCTCGCATTTCTCAGTACTATCTCAAAGAGGAGC
CTTTACTAATCTCGCACCTTCCCAAGTCGCGAGTGGGCTTACGCGCCGCCGCCACATAC
GGGATTGAGGCCAATGTCTAACCCTAACATTATCACCTCAGGCTAACAAATCCAGGACC
TCCTTTCTCGGATTTTGGACACACCATTCACATGGGAGTGGTCTCCTCTGCTTCTGATGC
TGATGTGCAACCGCCACCAGCCACCGCCACCAGAGGAACCGATGGTTAAGAGGAAACG
TGGACGGCCAAGAAAGTATGGAGAACCGATGGTTAGTAATAAGTCTAGGGACTCTTCTCC
AATGTCTGATCCTAATGAACCTAAACGGGCCAGAGGTGACCTCCTGGAAGTGAAGGAA
GCAACGCTTGCTAATCTTGGTGAGTGGATGAATACTTCAGCTGGACTTGCTTTTGCAAC
TCATGTGATCAGCATTGGAGCAGGAGAAGACATTGCTGCGAAAGTTTTGTCAATTTTACA
ACAAAGACCTCGGGCTCTTTGTATAATGTCAGGCACTGGAACCATTTCTTCAGTCACTCT
GTGCAAACCGGTTCAACCGATCGTCACTTAACATACGAGGGACCTTTTGGATTATAAG
TTTTGGTGGATCTTATTTGGTGAATGAAGAAGGTGGATCCAGAAGTGAACAGGCGGATT
GAGTGTCTCTCTTCTCGTCCCGATGGTAGTATTATTGCCGGTGGAGTTGACATGCTTAT
CGCAGCCAACTTGTTCAAGTGGTGGCATGTAGTTTTGTATACGGAGCAAGGGCAAAGAC

TCATAATAACAATAACAAGACCATCAGACAAGAAAAGGAACCAAATGAAGAGGACAACAA
TAGTGAAATGGAGACCACACCGGGTAGTGCAGCTGAACCAGCAGCATCTGCGGGTCAGCA
GACGCCACAGAACTTCTCTTCTCAGGGAATAAGGGGGTGGCCCGTTTCAGGCTCAGGCTC
TGGCAGATCATTGACATTTGCAGAAACCCACTCACTGATTTTGATTGACTCGTGGATG
ATATACACTATTAGTCTTTGAAGCAGCAGCATACAAAATGTGATTGCTGTACATATGTTA
TTGTAGATTTCTCTCTGGGAATGTTGAAATCAGACATTTAAGGATTGATACTAGATCTCT
CAGCTCCTTCTAACATTGTTAATGTAACAGAACCTCCCCTTTTCATGCTATTTGC
>G2151 Amino Acid Sequence (domain in AA coordinates:93-113, 124-144)
MDGREAMAFPGSHSQYYLQRGAFTNLAPSQVASGLHAPPPHTGLRPMSPNIHHPQANNP
GPPFSDFGHTIHMGVVSSASDADVQPPPPPPPEEPMVKRKRGRPRKYGEPMVSNKSRDS
SPMSDPNEPKRARGRPPTGRKQRLANLGEWMNTSAGLAFAPHVISIGAGEDIAAKVLSF
SQQRPRALCIMS GTGTISSVTLCCKPGSTDRHLTYEGPFEEISFGGSYLVNEEGGSRRTG
GLSVLSRPDGSIIAGGVMDLIAANLVQVVACSFVYGARAKTHNNNNKTIHQEKEPNEED
NNSEMETTPGSAAEPAASAGQQTPONFSSQIRGWPGSGSGSRSLDICRNPLTDFDLTR
G*

>G2154 (82..1317)

GCAAAAAGAAAAAATGAAAAAATCCCTAACTCTCTCTCTAGAAATCTTATTTTTG
TGCGTATCTCTCTAAAAAGGAATGGATCCTAACGAAAGCCACCATCACCACCAACAACAA
CAGCTCCATCACCTCCACCAACAGCAACAGCAGCAGCAGCAACGACTCACTTCT
CCTTACTTCCACCACCAACTACAGCACCATCACCACCTTCCAACCACCGTAGCAACCACC
GCTTCTACCGGAAACGCCGTTCCATCTTCCAACAATGGGCTTTTCCCTCCGAGCCTCAG
CCACAGCACCAGCCTAATGATGGGTCTCTCTCGCGGTGTACCTCATTCAGTTCCG
TCCTCGGCTGTGACGCGCGGATGGAGCCGGTAAAGAGGAAGAGGGGTCGACCAAGAAAG
TATGTGACGCCGGAACAAGCCCTAGCGGCTAAGAAATGGCGTCTTCTGCGAGTAGTTCTG
TCTGCTAAACAGAGCGAGAGCTTGCTGCTGTACCAGGTGGTACGGTATCGACTAATTCC
GGGTCTATCCAAGAAATCTCAGCTTGGTTCTGTGCGGAAAACTGGACAATGTTTACTCCG
CATATTGTTAATATAGCTCCTGGCGAGGATGTGGTCCAGAAAATTATGATGTTGCGAAAC
CAAAGCAAGCATGAATATGCGTTCTTCTGTCATCAGGCACTATCTAATGCATCCTTG
CGCCAACCGGCTCCATCAGGAGGCACTTACCATATGAGGGTCAATACGAGATTCTCTCA
CTATCTGGATCCTATATCCGAATGAACAAGGTGGTAAATCCGCGCGCCTTAGCGTTTCT
TTATCTGCTTCAGATGGTCAGATCATCGGTGGAGCGATTGGTAGCCATCTCACAGCTGCT
GGCCCGGTTTCAGGTGATTCTTGGTACGTTTTCAGCTTGATAGAAAGAGGATGCCGCCGGG
AGTGGTGGGAAAGGGGATGCTTCAAACAGTGGAAAGTCCGGTTAACTTCTCTGTAGCTCT
GGACAGTTGCTTGGCATGGGTTTCCCTCCTGGTATGGAATCTACGGGAAGAAATCCAATG
AGGGGAACAGCAGCAACATGATCATCATCATCAAGCCGGTTTGGGTGGACCTCAT
CATTTTCATGATGCAAGCGCGCAGGGGATACACATGACACATTCCAGGCCATCTGAATGG
CGCGGAGGAGGCAACAGCGGTCTGATGGCAGAGGCGGTGGCGGGTATGATTGTCAGGA
AGGATAGGACATGAGTCGTCGGAGAATGGAGATTACGAGCAGCAAAATACCGGATTAGCAG
AGCTTCCAGGAGAAGTGTGTAGAGTTTAGATCCCAAGTAGAGAAACAGAAGCGAGCAAA
GAATCTGAACTGAGAGAGGACTTATTAGACAGAGACTCGTCTGAAGGGTCTTTAATCATA
GAAAGAAGTTGCTGAGTGATTGCTTTTGTCTTCTTCTTGGTACGGTGTATTATATTAAAC
TCCACAACCTTTTTTTTATACTTTCAGTAACGATTCTCTTCACTTTCAATTTTCATTCTT
TTTTTTTATACTCTTTTTCTTTCTTATAATATTTTTTTTGGTTTTTCTTTCGTTTGTTA
CTAAAAAAGGAAATGCTCTTTTGTGAAATATATACACTTCGTTTTG

>G2154 Amino Acid Sequence (domain in AA coordinates:97-119)
MDPNESHHHHQQQLHHLHQQQQQQQQLTSPYFHHQLQHHLHLPPTTVATTASTGNAV
PSSNNGLFPPQPQPQHQPNDGSSSLAVYPHSPVSSAVTAPMEPVKRRGRPRKYVTPEQA
LAAKKLASSASSSSAKQRRELAAVTGGTVSTNSGSSKSKQLGSVKGTCQCFPHIVNIAP
GEDVVQKIMMFANQSKHELCLVLSASGTISNASLRQPAPSGGNLPYEGQYELSLSGSYIR
TEQGGKSGGLSVLSASDQIIGGAIGSHLTAAGPVQVILGTFQLDRKDAAGSGGKGD
SNSGSRSLTSPVSSGQLLGMGFPFGMESTGRNPMRGNDEQHDHHHQAGLGGPHHFMMPAP
QGIHMTSRPSEWRGGGNSGHDGRGGGGYDLSGRIGHESSENGDYEQQIPD*

>G2157 (306..1238)

TCTTTTGATTTTAACCTTTTTTTCAGTAGCAAGCCAAAAAAGAGCAAAAGAAGTT
CCTTTTATGATAAAGGTATGATGATAGCAACAAATGATACCCCATGTCTTGTGTGTCT
GCTTCATGCAACATGTTGGTTTGGATTGTTAATCTAAAAGTTTAAGATAAGGTTTTTCG
GATTCTCTTCTGTCTTGTAAATAGTTTCTTGTGCGAGAGCCATCAACACCAACTTCAACA

AAAAAACAAGAAAAAGAAAAAGATTCTCTTTCTCGTTTATTTCCATTAGAGAAGAAAA
AAAGAATGGCGAATCCTTGGTGGGTAGGGAATGTTGCGATCGGTGGAGTTGAGAGTCCAG
TGACGTCATCAGCTCCTTCTTTGCACCACAGAAACAGTAACAACAACACCCACCGACTA
TGACTCGTTCGGATCCAAGATTGGACCATGACTTCACCACCAACAACAGTGGAAGCCCTA
ATACCCAGACTCAGAGCCAAGAAGAACAAGACAGCAGAGACGAGCAACCAGCTGTTGAAC
CCGGATCCGGATCCGGGTCTACGGGTCTGTCCTAGAGGTAGACCTCCTGGTTCCAAGA
ACAAACCAAGAGTCCAGTTGTTGTTACCAAAGAAAGCCCTAACTCTCTCCAGAGCCATG
TTCTTGAGATTGCTACGGGAGCTGACGTGGCGGAAAGCTTAAACGCCCTTGTCTCGTAGAC
GCGGCCGCGGCGTTTCGGTGTGAGCGGTAGTGGTTTGGTTACTAATGTTACTCTGCGTC
AGCCTGCTGCATCCGGTGGAGTTGTTAGTTTACGTGGTCAGTTTGAGATCTTGTCTATGT
GTGGGGCTTTTCTTCTACGTCTGGCTCTCCTGCTGCAGCCGCTGGTTTAAACATTACT
TAGCTGGAGCTCAAGGTCAAGTTGTGGGAGGTGGAGTTGCTGGCCCGCTTATTGCCTCTG
GACCCGTTATTGTGATAGCTGCTACGTTTGAATGCCACTTATGAGAGGTTACCGATTG
AGGAAGAACAACAGCAAGAGCAGCCGCTTCAACTAGAAGATGGGAAGAAGCAGAAAGAAG
AGAATGATGATAACAGAGATGGGAATAACGGAACGAAGGATCGATGCAGCCGCCGATGT
ATAATATGCCTCCTAATTTTATCCCAATGGTCATCAAATGGCTCAACACGACGTGTATT
GGGGTGGTCTCCGCCCTGCTCTCCTTCGTATTGATTAGTTAGATAGGCGGTGGTTG
GTGCGTTCTTTTACTGGAATGATTATATTTTCCATTAGGATGGTTAGGCTTTTGTATT
TAAAGCTATCAAGTTTCTTTTACGGATAATTCGGATGACAATTAGCTAGTGTGT
GTTTGTGTTTGTGGCGGCTTTTCTGACTTGACTATTTTGATCGCGGATAGCTTTGTA
TGAAAGTGAATTGATTGTAGAATCGTCTTTGAATTTGATGTTGAAAAAACCA
>G2157 Amino Acid Sequence (domain in AA coordinates: 82-102, 164-107)
MANPWWVGNVAIGGVESPTSSAPSLHHRNSNNNPPTMTRSDPRLDHDFTTNSGSPNT
QTQSQEEQNSRDEQPAVEPGSGSGTGRPRGRPPGSKNPKSPVVVTKEPNSLQSHVL
BIATGADVAESLNAFARRRGRVSVLSGSLVTNVTLRQPAASGGVSVLRGQFEILSMCG
AFLPTSGSPAAAAGLTIYLAGAQGVVGGGVAGPLIASGPVIVIAATFCNATYERLPIEE
EQQQEQPLQLEDGKKQKEENDDNESGNNGNEGSMQPPMYNMPPNFIPNGHQMAQHDVYWG
GPPPRAPPSY*
>G2181 (1..1005)
ATGATGCTTGCGGTGGAAGATGTGTTAAGCGAACTCGCCGGAGAAGAAAGGAACGAGAGA
GGATTGCCACCTGGCTTCCGGTTTACCCGACGGACGAAGAGCTCATTACCTTCTACTTA
GCTTCCAAAATCTTCCATGGTGGTCTCTCCGGCATTCACTTCCGAAGTTGATCTCAAC
CGCTGTGAACCTTGGGAGCTACAGAAATGGCGAAGATGGGAGAGAGAGAGTGGTACTTT
TATAGTCTAAGGGACAGGAAATATCCGACAGGTTTGAGGACTAACAGAGCAACTACTGCT
GGTACTTGAAAGCTACCGGCAAAGATAAGGAAGTCTTCTCCGGCGGAGGAGGACAGCTT
GTTGGGATGAAGAAGACGTGTGGTGTCTACAAAGGTAGGGCTCCACGTGGCCTCAAGACT
AAGTGGGTGCATGCATGAGTATCGCCTCGAAAACGACCATTCACACCGCCACACGTGTAAG
GAGGAATGGGTGATTGTCAGAGTGTCAATAAAACAGGAGACAGAAAAATGTTGGATTA
ATCCATAACCAATCAGTACCTTCATAACCATTCACTCTCAACAACACATCATCATCAT
CATGAAGCCTTACCTTTGCTTATAGAACCTTCCAACAAAACCTAACCAACTTCCCATCA
CTACTCTACGATGATCCACCAAACTACAATAATAACAACCTTCTTCATGGATCATCA
GGCCACAACATCGACGAGCTCAAAGCCTTAATCAACCCTGTCTCTCAGCTCAACGGT
ATCATCTTTCTTCAGGGAAACAACAACGACGAAGACGACTTCGACTTTAACCTCGGC
GTGAAAACAGAGCAGTCTTGAACGGTAACGAAATTGACGTACGAGATTACTTGGAGAAC
CCTCTGTTTCAGGAAGCGAGTTATGGTCTGTTGGGTTTTTCGCTCTCTCTGGACCTCTT
CACATGCTACTAGATTCTCCATGTCTTTAGGATTCCAGCTGTAG
>G2181 Amino Acid Sequence (conserved domain in AA coordinates:22-169)
MMLAVEDVLSSELAGEERNERGLPPGFRFHPTDEELITFYLASKIFHGGLSGIHISEVDLN
RCEPWELPEMAKMGEREWFYFSLRDRKYPTGLRNRATTAGYWKATGKDKEVFSGGGGQL
VGMKKTIVFYKGRAPRGLKTKWVMHEYRLNENDHSHRHTCKEEWICRVFNKTGDRKNVGL
IHNQISYLNHNSLSTTHHHHHEALPLLI EPSNKTLTNFP SLLYDPHQNYNNNNFLHGSS
GHNIDELKALINPVVSQNLNGIIFPSGNNNNDEDDFDNLGVKTEQSSNGNEIDVRDYLEN
PLFQEASYGLLGFSSSPGPLHMLLDSPCPLGFQL*
>G221 (115..795)
CTCTCTTATTCTCTCACTCTTTTTTTTTTATATTCTCTCTCTCTAAATCTATAAAATAT
ATTTAAAAACTTGATCGTATATAATAAAGTAAATAAAGAATAATAACAAAAAATGGAG
AAAAGAGGAGGAGGAAGTAGTGGAGGTTCCGGATCATCAGCAGAAGCAGAAGTGAGAAAA

GGACCATGGACGATGGAAGAAGATCTTATTCTTATCAACTATATCGCCAACACGGCGAT
GGTGTGTTGGAATCTCTCGCCAAATCTGCAGGTCTAAACGAACCGGGAAAAGTTGCCGG
CTCCGGTGGCTGAAGTATCTCCGCCCGACGTACGACGGGAAACATCACTCCAGAAGAG
CAACTTATCATCATGGAACCTCATGCTAAGTGGGGAAACAGGTGGTCGAAAATCGCCAAA
CATCTTCAGGAAGAACGGACAACGAGATCAAAAATTTCTGTAGGACAAGAATTCAAAAA
TACATCAAGCAATCGGATGTAAACAACATCGTCCGTTGGATCTCATCATAGCTCAGAG
ATCAACGATCAAGCTGCAAGCAGCTCGAGCCATAATGTCTTTGTACACAAGATCAAGCG
ATGGAGACTTATCTCTACACCGACATCATATCAACATACCAATATGGAATTCAACTAT
GGTAACTATTCCGCCCGCGGACGTGACGGCAACCGTGGATTATCCAGTACCGATGACCGTT
GATGATCAAACCGGTGAAAATATTGGGGCATGGATGATATTGGTCATCAATGCATTTA
TTGAATGGTAATTGATTGATCGGTGGACAAAACATGGAATATTAATTGAGTATTATATAT
GATTTTATAGGAGTACTATTATTAGTACGTGACATGTATATGTTTTGCCTCGTTGTAGAG
GTTTGGGGTTATAATTAATATATAATGTTATCTAATATGCAACCTTGATACATATTTGGA
TCTTTATTGAACCCATGTTATACATAAATAAAATTGTTGAAGGGGTCATAAAAAAAAAA
AAAAAAAAAAAAA

>G221 Amino Acid Sequence (domain in AA coordinates: 21-125)
MEKRGSSGSSGSSAEVRKGPWTMEEDLILINYIANHGDGVWNSLAKSAGLKRTGKS
CRLRLWNLRLPDRVRRGNITPEEQLIIMELHAKWGNRWSKIAKHLPGRTDNEIKNFCRTRI
QKYIKQSDVTTTTSSVSHHSSEINDQAASTSSHNVFCTQDQAMETYSPTPTSQHTNMEF
NYGNYSAAAVTATVDYPVPMTVDDQTGENYWGMDIWSMHLNNGN*

>G2290 (119..982)

TTCTTTCTTTCTTTCTTTCTTCTTCCAATCAAGAACAAACCCCTAGCTCCTCTCTTTTCTC
TCTCTACCTCTCTTTCTCTATCTTCTTATCACTACTTCTCTCGCCGATCAATCATCAT
GAACGATCCTGATAATCCCGATCTGAGCAACGACGACTCTGCTTGGAGAGAAGTCAACT
CAGAGCTCAAGATTCTGACTTCTTCGACCGAGACACTTCCAATATCCTCTCTGACTTCGG
TTGGAACCTCCACCCTCCTCCGATCATCCTCACAGTCTCAGATTGCACTCCGATTTAAC
ACAAACCACCGGAGTCAAACCTACCACCGTCACTTCTTCTTGTTCCTCATCCGCCGCCGT
TTCCGTTGCCGTTACTCTACTAATAATAATCCCTCAGCTACCTCAAGTTCAAGTGAAGA
TCCGCCCGAGAAGTCAACCGCTCCGCCGAGAAAACACCACCACCGGAGACACCAAGTGA
GGAGAAGAAGAAGGCTCAAAAGCGAATTCCGCAACCAAGATTTCGATTTCATGACCAAGAG
TGATGTGGATAATCTTGAAGATGGATATCGATGGCGTAAATATGGACAAAAGCCGTCAA
GAATAGCCCATTTCCCAAGGAGCTACTATAGATGCACAAACAGCAGATGCACGGTGAAGAA
GAGAGTAGAAGCTTTCATCAGATGATCCATCGATAGTGATCACAACATACGAAGGACAACA
TTGCCATCAAACCATTTGGATTCCCTCGTGGTGAATCCTCACTGCACACGACCCACATAG
CTTCACTTCTCATCATCATCTCCCTCCTCCATTACCAAATCCTTATTATTACCAAGAACT
CCTTCATCAACTTCACAGAGACAATAATGCTCCTTCACCGCGGTTACCCCGACCTACTAC
TGAAGATACACCTGCCGTGTCTACTCCATCAGAGGAAGGCTTACTTGGTGATATTGTACC
TCAAACATATGCGCAACCCCTTGAGGTAAGCTTGGTACGTAGCAATAGCTAAGGAGGTGCTA
ACTCATTATATATAGAAGATATTGCAGACAGAAATATGCGCAGGGAGGGTATAACAATAT
GGCGTTGTAACAATGGATCTATATATTACCTCATTTGTTGATCAATAGCACACCACCGGTA
CGTTTGCAATTTCTTCATGTATATTCTTGTATATATGTAGTTATATATCCAGGTATAA
TTTTGATGTAACACAACATTAATCTTAATCGTGGATCCATCCACATTTGATGCATGTAT
GTGCACTTAAGAAAAAGAACATGGAGGAAATAACGTATTTTTTTTATTATTCT

>G2290 Amino Acid Sequence (conserved domain in AA coordinates:147-205)

MNDPDNPDLSNDDSAWRELTLTAQDSDFDRDTSNILSDFGWNLHSSDHPHSLRFDSDL
TQTTGVKPTVTSSCSSSAVSVAVTSTNNNPSATSSSSSDPAENSTASAETKPPPETPV
KEKKKAQKRIRQPRFAFMKSDVDNLEDGYRWRKYGQKAVKNSPFPRSYRCTNSRCTVK
KRVERSSDDPSIVFTTYEQGHCHQTIGFPRGGILTAHDPHSFTSHHHLPPLPNPYYYQE
LLHQLHRDNNAPSRLPRPTTEDTPAVSTPSEEGLLGDIVPQTMRNP*

>G2299 (231..941)

GCCAAAATTTTACCAACATTTTCTTCTCATATCAAAGTTTCTCTCTCATTTCTTCAT
CACACTTCACTGCCCTGTTTTTTTTCTCATTTTGAATAGTTCTCAAACCTATATATTTT
TCCCCTGAAGCCTAGCTATTTCTTTTATTGTCATTAATCTCGGATCCGAATCGAAAA
AAGCAATCAGAATAATAGACTTGTACGATACTGTGCTTAAGCTAACACAATGGCAGAGG
AATACTACAGCCTCCGCTCGGAGAGAGTAACTCAGCTTCTTGTCCCTAACTCGGAGTCTG
ACTCAGTGAGTGACAAAAGCAAAGCTGAGCAAAGCGAGAAGAAGACTAAACGTGGGAGAG
ACTCCGGTAACACCCTGTTTATCGCGGAGTAAGGATGAGGAAGTGGGGAAAATGGGTGT

CGGAGATTTCGTGAGCCGAGGAAGAAATCACGTATTTGGCTGGGAACCTTTCCCGACGCCGG
AGATGGCGCGCGCTGCACACGACGTGGCGGCTCTGAGCATTAAAGGAACGGCCGCTATAC
TAAACTTCCCTGAACTCGCTGACTCATTCCCTCGACCCGTTTCATTAAGCCCTCGAGACA
TTCAGACAGCAGCTCTTAAAGCAGCTCACATGGAACCGACGACGTCGTTTTTCATCTTCCA
CGTCTTCGTCGTCGTCTTTGTCTTCTACGTCTTCGCTCGAGTCTCTTGTGTTGGTGATGG
ACCTCTCGAGGACTGAGTCGGAGGAGCTCGGTGAGATTGTGGAGCTTCCAAGTCTCGGGG
CGAGTTACGACGTCGACTCGGCTAACCTTGGGAACGAGTTTGTCTTCTATGACTCAGTTG
ACTACTGTTTTATATCCGCCGCCGTGGGGACAGTCGTCCGAAGATAACTATGGTCACGGAA
TTAGCCCTAATTTTGGCCATGGCTTGTCTATGGGATCTCTAACAGTTTATTTTGTATCATT
ACCATAATGTTTTGTTTAAACAGTTTATTTTGTATCATTGCCATAATGTTTTGTTTAAAT
CACGTTTTTAAACCCCTTTGCTGTTTTTGTTTTTTTTTTGAGTTTTT

>G2299 Amino Acid Sequence (conserved domain in AA coordinates:48-115)

MAEYYSLRSERVQLLPNSESDSVSDKSKAEQSEKKTKRGRDSGKHPVYRGVVRMRNWG
KMWSEIREPRKKSRIWLGFPTPEMAARAHDAALSIKGTAAILNFPPELADSFPRPVSL
PRDIQTAALKAAHMEPTTSFSSSTSSSSSLSTSSLESVLVMDLSRTESEELGEIVELP
SLGASYDVDSANLGNFVFYDSVDYCLYPPPWGQSSSEDNYGHGISPNFGHGLSWDL*

>G2340 (274..1275)

ATACAAACTCCCTCTTCTCTATCTTCTTCATCTTAAAGAAAAAATAAGAGATATTCGTA
AAGAGAGAACAACAAATTTTCACTTACGAAAAGCTAGCAAAGTCGAGTATCGAGGAATAA
CAGAATAAGACGTATCTATCCTTGCCCTTAATGTTCTTACCAAAGATCTAGTCCTTTCTT
TGATGATCGATCCATCACAAAGCCACAACAACAACACTACATCTCTTCTCTATCTCT
AGCTTCTATTTTAAATACATTCAAGAATCAAGAATGGTACGGACGCCGTGTTGTAGAGCA
GAAGGGTTGAAGAAAGGAGCATGGACTCAAGAAGAAGACCAAAGCTTATCGCCTATGTT
CAACGACATGGTGAAGGCGGTTGGCGAACCCCTCCGGACAAAGCTGGACTCAAAAGATGT
GGCAAAGCTGCAGATTGAGATGGGCGAATTACTTAAAGACCTGACATTAAACGTGGAGAG
TTAGCCAAGACGAGGAAGATTCCATCATCAACCTCCACGCCATTATGGCAACAAATGG
TCGGCCATAGCTCGTAAATAACCAAGAAGACAGACAATGAGATCAAGAACCATTGGAAC
ACTCACATCAAGAAATGTCTGGTCAAGAAAGGTATTGATCCGTTGACCCACAAATCCCTT
CTCGATGGAGCCGGTAAATCATCTGACCATTCCGCGCATCCCGAGAAAAGCAGCGTTCAT
GACGACAAAGATGATCAGAAATTCAAATAACAAAAGTTGTCAAGATCATCATCAGCTCGG
TTTTTGAACAGAGTAGCAAAACAGATTCCGTCATAGAATCAACCACAATGTTCTGTCTGAT
ATTATTGGAAGTAATGGCCTACTTACTAGTCACACTACTCCAACACAAAGTGTTCAGAA
GGTGAGAGGTCAACGAGTTCCTCTCCACACATACCTCTTCAATCTCCCATCAACCGT
AGCATAACCGTTGATGCAACATCTCTATCCTCATCCACGTTCTCTGACTCCCCCGACCCG
TGTTTATACGAGGAAATAGTCGGTGACATTGAAGATATGACGAGATTTTCATCAAGATGT
TTGAGTCATGTTTTATCTCATGAAGATTTATTTGATGTCGTTGAGTCTGTTTGGAGAAT
ACTTCATTTCATGAGGGAAATTAACAATGATCTTTCAAGAGGATAAAATCGAGACGACGTCG
TTTAATGATAGCTACGTGACGCCGATCAATGAAGTTGATGACTCCTGTGAAGGGATTGAC
AATTATTTTGGATGAGTTATATTGATGATGATGAAAATTTGCATTTGGCATGTAAATCAA
TTAGAGTTTGATTTGCTATGGTGTTTTTAGTTTGTGTGTGTAGTGTGTTTCGACCGTCAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAA

>G2340 Amino Acid Sequence (domain in AA coordinates:14-120)

MVRTPCCRAEGLKKGAWTQEEDQKLIAYVQRHGEQWRTLPDKAGLKRCKGKSCRLRWANY
LRPDIKRGEFSQDEEDSIINLHAIHGNKWSAIARKIPRRIDNEIKNHNTHIKKCLVKKG
IDPLTHKSLLDGAGKSSDHSAPHEKSSVHDDKDDQNSNNKLSGSSSARFLNRVANRFGH
RINHNVLSDIIGSNGLTSHHTPTTSVSEGERSTSSSSTHTSSNLPINRSITVDATSLSS
STFSDSPDCLYEEIVGDIEDMTRFSSRCLSHVLSHEDLLMSVESCLENTSFMREITMIF
QEDKIETTSFNDSYVTPINEVDDSCGIDNYFG*

>G2346 (1..1011)

ATGGAGTTGTTAATGTGTTCCGGGTGAGCCGAGTCAGGTGGTTCTTCTTCCACCGAGTCT
TCTTCACTCAGTGGTGGACTCAGGTTTGGTCAGAAGATCTACTTCGAGGATGGATCCGGA
TCCAGAAGCAAGAACCGGTCATACCGTTTCGTAAGTCGTCTACCACGGCGAGGTGCCAA
GTGGAAGGTTGTAGAAATGGATCTAAGCAATGTTAAAGCTTATTACTCGAGACACAAAGTT
TGTTGCATTCACTCTAAATCATCTAAAGTCATTGTCTCTGGTCTTCATCAAAGGTTTGT
CAACAATGTAGCAGGTTTACCAGCTTCTGAGTTTGACTTGGAGAAAAGAAGTTGTCGC
AGAAGACTCGCTTGTACATAACGAACGACGAAGAAAACCAACCCACAACCGGCTCTTTTC
ACTTCTCATTACTCTGAATCGCTCCATCTCTTTACGGAACCCCAATGCTGCAATGATT

AAAAGCGTTTTGGGAGATCCTACTGCGTGGTCAACCGCAAGATCAGTGATGCAGCGGCCT
GGACCGTGGCAGATTAATCCAGTTAGGGAAACCCATCCACACATGAATGTTTTATCACAT
GGAAGCTCAAGCTTTACTACATGTCCAGAGATGATAAACAACAATAGCACAGATTCAAGC
TGTGCTCTCTCTCTCTGTCAAACCTCATACCCAATTCATCAGCAGCAACTTCAGACACCA
ACAAATACATGGCGACCATCTTCTGGTTTCGACTCGATGATCTCATTCTCCGATAAGGTT
ACAATGGCTCAGCCACCGCCCATTTCAACCCATCAGCCGCCCATCTCAACACATCAGCAG
TACCTCAGCCAAACTTGGGAAGTCATCGCGGGCGAAAAGAGCAATTACATTATATGTCT
CCTGTGAGTCAAATCTCGGAGCCAGCAGATTTCAGATAAGCAATGGCAGTGTGTCGCC
TATTCTCCTCCGTCCTTACTATCTCTTGTGTGCTACTTGCGCCGCTATAG

>G2346 Amino Acid Sequence (domain in AA coordinates: 59-135)
MELLMCSGQAESGSSSTESSLSGGLRFGQKIYFEDGSGSRSKNRVNTVRKSSTTARCO
VEGCRMDLSNVKAYYSRHKVCCIHSKSSKVIIVSGLHQRFCQQCSRHFHQLSEFDLEKRSCR
RRLACHNERRRRKPQPTTALFTSHYSRIAPSLYGNPNAAMIKSVLGDPTAWSTARSVMQRP
GPWQINPVRETHPHMNVLSHGSSSFTTCPEMINNNSTDSSCALSLNSYPIHQQLQTP
TNTWRPSSGFDMSISFSDKVTMAQPPPISHTQPPISHTQQLSQTWEVIAGEKSNSHYMS
PVSQISEPADFQISNGSVSPYSPPSLLSLVCLRLPL*

>G237 (1..852)

ATGGCGAAGACGAAATATGGAGAGAGACATAGGAAAGGGTTATGGTCACCTGAAGAAGAC
GAGAAGCTAAGGAGCTTCACTCTCTTATGGCCATTCTTGCTGGACCACTGTTCCCATC
AAAGCTGGGTACAAAGGAATGGGAAGAGCTGCAGATTAAAGATGGATTAAATTACCTAAGA
CCAGGGTTAAAGAGGGATATGATTAGTGCAGAAGAAGAAGAGACTATCTTGACGTTTCAT
TCTCCCTTGGGTAACAAGTGGTCGCAATAGCTAAATTTTACCGGAAGAACAGACAAT
GAGATAAAGAACTATTGGCACTCTCATTGAAAAAGAAATGGCTCAAGTCTCAGAGCTTA
CAAGATGCAAAATCTATTTCCCTCTCTCGTCTTCATCATCATCACTTGTGTGCTTGTGGA
GAAAGAAATCCGGAACCTTGATCTCGAATCACGTGTTCTCCCTCCAGAGACTTCTAGAG
AACAAATCTTCATCTCCCTCACAAGAAAGCAACGGAATAACAGCCATCAATGTTCTTCT
GCTCCTGAGATTCCAAGGCTTTTCTCTCTGAATGGCTTTCTTCTCATATCCCCACACC
GATTATTCTCTGAGTTTACCGACTCTAAGCACAGTCAAGCTCCAATGTGGAAGAGACT
CTCTCAGCTTATGAAGAAATGGGTGATGTTGATCAGTTCATTACAACGAAATGATGATC
AACAAAGCAACTGGACTCTTAACGACATTGTGTTTGGTTCCAAATGTAAGAAGCAGGAG
CATCATATTTATAGAGAGGCTTCAGATTGTAATTTCTCTGCTGAATTTCTTCTCCACCA
ACAACGACGTAAATTGCGTTTATTGTAATGTAAATCAAATTTCTAAGGCAAAACCGGAAA
AAAAAAAAAAAAAAAAAAAAA

>G237 Amino Acid Sequence (domain in AA coordinates: 11-113)
MAKTKYGERHRKGLWSPEDEKLRSFILSYGHSCWTTVPIKAGLQRNGKSCRLRWINYLR
PGLKRDMSIAEEETILTFSPLGNKWSQIAKFLPGRDNEIKNYWHSLSLKKKWLKSQSL
QDAKSISPPSSSSSSSLVACGERNPETLISNHVFSLQRLLENKSSSPSQESNGNNSHQCSS
APEIPRLFFSEWLSSSYPHTDYSSEFTDSKHSQAPNVEETLSAYEEMGDVDQPHYNEMMI
NNSNWTLNDIVFGSKCKQEHIIYREASDCNSSAEFFSPPTT*

>G2373 (48..1199)

GCAAAATCCTCAGATCGTCTTACCTTCTCCGAATCGATCGATTTTTTCATGGAGGACGACG
ACGAGATTCAATCCATCTCCGGGAGATTCTTCCCTTTACCACAAGCTCCTCCTT
CTCCGCCGATTTTGCCAACAAACGACGTGACGGTGGCCGTCGTGAAGAAACCACAACCGG
GGCTTTCTTCTCAATCTCCGTCCATGAACGCTTTAGCGTTAGTGGTTCATACTCCTTCTG
TAACCGGTGGTGGTGGTAGCGGAAACAGAAACGACGAGGAGGAGGAGGAAGCGGTG
GTGGTGGAGGAGGAAGAGATGATTGTTGGAGCGAAGAAGCTACAAAGGTTCTAATCGAAG
CTTGGGGAGATCGATTCTCTGAACGAGTAAAGGAACCTTGAAGCAACAACATTGGAAAG
AAGTAGCTGAGATTGTGAACAAGAGTCGTCAATGCAAAATACCCTAAAACGTGATATTCAAT
GTAAGAACAGAAATTGATACGGTGAAGAAGAAGTATAAGCAAGAGAAAGCTAAGATTGCTT
CTGGTGATGGACCTAGTAAATGGGTTTTCTTCAAGAAGCTTGAGAGTTTGATTGGTGGTA
CTACAACATTCAATGCTTCTTCAAAAGCTTCAGAGAAGGCTCCTATGGGAGGAGCTCTTG
GGAATAGCCGTTTCGAGTATGTTTAAACGGCAAACTAAAGGTAATCAGATTGTGCAGCAAC
ACAAGAGAAGAGAGGCTCTGATTGATGCGGTGGCATTTTAGGAAACGTAGTGCTTCTG
AGACTGAGTCTGAGTCTGATCTCTGAACCTGAGGCTTCTCTGAGGAATCTGCTGAGAGTC
TCCACCTTTGCAACCGATTCAACCGCTTTCGTTTTCATATGCCAAAGCGGTTGAAGGTGG
ATAAGAGTGGAGGTGGAGGAGTGGAGTGGAGATGTGGCGAGGGCGATACTTGGATTTA
CGGAAGCTTATGAGAAGGCGGAAACTGCTAAGCTTAAGTTAATGGCGGAACCTGGAAAAGG

AGAGGATGAAATTTGCTAAAGAGATGGAGTTGCAGAGAATGCAGTTCTTGAAACTCAAT
TGGAGATAACACAGAACAAATCAAGAAGAGGAAGAGAGGAGCAGCGAGGAGAAAGGA
GGATCGTTGATGATGATGATGATGATCGCAATGGCAAGAATAACGGCAATGTAAGTAGCTGAC
AATTGAACACACAAATGTTCTATGATATTTGCTATGATAAGCTGGATTTTAGGTTTTGA
TGG

>G2373 Amino Acid Sequence (domain in AA coordinates:290-350)
MEDDDEIQSIPSPGSSLSPPQAPPSPILPTNDVTVAVVKKPQPLSSQSPSMNALALVV
HTPSVTGGGSGNRNRRGGGGGGGGGGGRDDCWSEETKVLIEAWGDRFSEPGKGLKQ
QHWKEVAEIVNKSQCKYKPKTDIQCKNRIDTVKKKYKQEKAKIASGDGSPKWFVFKKLES
LIGGTTTTFIASSKASEKAPMGALGNSRSMFKRQTKGNQIVQQQEKRGSDSMRWHFRK
RSASETESESDPEPEASPEESAEPLPPIQPLSFHMPKRLKVDKSGGGSGVGDVARA
ILGFTAEYEAETAKLKLMAELEKERMKFAKEMELQRMQFLKTQLEITQNNQEEERSRQ
RGERRIVDDDDDRNGKNNGNVSS*

>G2376 (39..1370)

CACGAGCTTCTGACTCAGATCCGGCGATATCGAATTCATGGAGGACGATGAAGACATCC
GATCTCAGGGTTCCGATTACCTGATCCGTCTTCTCCCCCGCGGGACGAATCACGG
TTACGGTGGCTTCGGCAGGTCCGCCTTCTTATTCTCTGACTCCTCCGGGTAATTCGTGCG
AGAAGGATCCGGATGCGTTGGCTCTGGCGCTGCTTCCGATTACAGCCAGCGGTGGAGGGA
ATAACAGCAGTGGGAGACCAACCGGCGGCGGGAGGGAGGATTGTTGGAGCGAAGCAG
CTACGGCTGTGTTGATTGATGCGTGGGGTGAGAGATACTTGGAGCTTAGCAGAGGGAATC
TGAAGCAGAAGCACTGGAAGAGGTGGCTGAGATTGTGAGCAGCAGAGAGGATTACGGTA
AAATTCCCAAACTGATATACAGTGTAAGAATAGGATCGATACGGTGAAGAAGAAGTATA
AACAAGAGAAGGTGAGAATCGCTAACGGCGGTGGCCGTAGCAGATGGGTGTTCTTCGACA
AGCTTGACCGTCTGATTGGATCAACGGCGAAGATCCCGACGGCAACTTCTGGAGTCAGCG
GTCCTGTGCGGAGGATTGCATAAGATTCTATGGGTATTCCAATGGGAAGTCGTTTCAATC
TGTACCATCAGCAAGCTAAGGCTGCAACACCGCCTTTCAATAATCTTGACCGGTTAATTG
GAGCTACGGCTAGAGTCTCAGCTGCTTCTTTCGGTGGCAGTGGTGGAGGAGCGGAGGAG
GATCTGTCAATGTACCTATGGGAATTCGATGAGTAGCCGTTTCACTCCGTTTGGACAGC
AAGGAGGACTCTGCCACAGCAAGGTAGGACACTGCCACAGCAACAGCAGCAAGGGATGA
TGGTGAAAAGGTGTAGTGAGTCAACACGCTGGCGTTTCAGGAAGAGGAACGCTTCTGATT
CAGACTCGGAATCTGAAGCAGCAATGTGATGATTCCGGTGACAGTTTACCACCTCCTC
CTCTGTGCAAGAGGATGAAGACGGAGGAGAAGAAGAAGCAAGATGGTGATGGAGTGGGA
ACAAATGGAGGGAGCTGACTCGGGCAATCATGAGATTCCGTGAAGCTTATGAGCAACAG
AGAATGCGAACTGCAACAGGTGGTTGAGATGGAGAAAGAGAGGATGAAGTTCTTGAAGG
AGCTTGAGTTGCAGAGAATGCAGTTCTTTGTGAAGACTCAATTGGAGATATCACAACTTA
AGCAGCAACATGGGAGGAGAATGGGAAACACCAGTAATGATCATCATCAGCCGCAAGA
ACAACATCAATGCGATTGTCAACAACAACGATTGCGTAATAACTAGAATTTAGTGA
TGCAGTGTGCTAATTGATATATTTTAGATTTGAG

>G2376 Amino Acid Sequence (domain in AA coordinates:79-178, 336-408)

MEDDEDIRSQSDSPDPSSPPAGRITVTVASAGPPSYSLTPPGNSSQKDPDALALALLP
IQASGGGNNSSGRPTGGGREDCEAATAVLIDAWGERYLELSRGNLQKHWKEVAEIV
SSREDYGIKPKTDIQCKNRIDTVKKKYKQEKVRIANGGGRSRWVFFDKLDRIGSTAKIP
TATSGVSGPVGGLHKIPMGIPMSRSNLYHQAKAATPPFNNLDRIGATARVSAASFSG
SGGGGGGGSVNVPMPGIPMSRSAPFGQQGRTLPQQGRTLPQQQQGMMVKRCSESKRWF
RKRNASDSDSEEAAMSDDSGDSLPPPLSKRMKTEKKKQDGDGVGNKWRELTRAIMRF
GEAYEQTENAKLQQVVEKERMKFLKELELQRMQFFVKTQLEISQLKQGHRRMGNTSN
DHHHSRKNINAIIVNNNDLGN*

>G24 (194..724)

CGGACGGTGGGCAAATATTAATAAAAAAGTGTGGTGAATTCTCAATCTTTGTCTTCT
TTCGTCTCTCTTTAAACTCCTCCGTCCCTCCTTATTATGTAACCGTCTCGCCGTCAA
TTTTCAAATCTCTCCCTCCGTTCAAAACCCAGATCGAAATTTATGGTTTTGTAATTTT
TTTACCGCGGTTATGGAGACGGAAGCGGCGGTGACAGCGAGGTTACGGCGGACGAT
GGGGATTGGGACGAGGAAGAGAGATCTGAAACCGTATAAAGGAATACGAATGAGGAAATG
GGGGAATGGGTGGCGGAGATACGGGAACCGAATAAGAGATCAAGGATCTGGTTAGGTTT
TTATGCGACGCTGAAGCGGCGGCGAGAGCTTACGACACTGCTGTTTTTACCTCCGTGG
TCCTTCAGCGAGGCTTAATTTTCCGGAGCTTTTGGCTGGACTTACTGTTTCTAACGGCGG
AGGAAGAGGTGGTGATTATCGGCGCGTATATAGGAGAAAAGCGGCGGAGGTTGGTGC

TCAGGTTGATGCGCTTGGAGCGACGGTGGTTGTGAATACCGGCGGCGAGAATCGCGGTGA
TTACGAGAAGATTGAGAATTGTCGTAAGAGCGGTAACGGGTCATTGGAACGGGTCGATTT
GAATAAATTACCCGACCCGAAAATTTCGGATGGTGATGATGACGAATGTGTGAAAAGAAG
ATAGAAAAATAAAAAGTAGTTGTAGAAGGAGAGACGAGAATGTTTGTCTTTAAGATGCG
CTGTTGCCGCTAACATGCGCTTTTCGATTTTAGTGTTAAACATGCGCCTCCATTGTTTTTG
GGTTTTGTTTTCGTCGTCGATAATCAAAGATTTTAAACACAATTCTCAAATTTTTCACT
TGTACAAACTAGATTTGCATGATCTTTGTATTAACGAATAACGATTAAGTCCTAAA

>G24 Amino Acid Sequence (domain in AA coordinates: 25-93)

METEAAVTATVTAATMGIGTRKRDLPYKGI RMRKWKWVAEIREPNKRSRIWLGSYATP
EAAARAYDTAVFYLRGPSARLNFPELLAGLTVSNGGGRGDLAAYIRRKAAEVGAQVDA
LGATVVVNTGGENRGDYEKIENCRKSGNSLERVDLNLKLPDPENSDGDDDECVKRR*

>G2424 (1..999)

ATGAGGATGGAGATGGTGCATGCTGACGTGGCGTCTCTCTCCATAACACCTTGCTTCCCCG
TCTTCTTTGTCTTCGTCCTCACAATCATCACTATAACCAACAACAACATTGTATCATGTGCG
GAAGATCAACACCATTTCGATGGATCAGACCACTTCATCGGACTACTTCTCTTTAAATATC
GACAATGCTCAACATCTCCGTAGCTACTACACAAGTCATAGAGAAGAAGACATGAACCCCT
AATCTAAGTGATTACAGTAATTGCAACAAGAAAGACACAACAGTCTATAGAAGCTGTGGA
CACTCGTCAAAAGCTTCGGTGTCTAGAGGACATTGGAGACCAGCTGAAGATACTAAGCTC
AAAGAACTAGTCGCGCTCTACGGTCCACAAAACCTGGAACCTCATAGCTGAGAAGCTCCAA
GGAAGATCCGGGAAAAGCTGTAGGCTTCGATGGTTTAAACCAACTAGACCCAAGGATAAAT
AGAAGAGCCTTCACTGAGGAAGAAGAAGAGAGGCTAATGCAAGCTCATAGGCTTTATGGT
AACAAATGGGCGATGATAGCGAGGCTTTTCCCTGGTAGGACTGATAATTCTGTGAAGAAC
CATTGGCATGTTATAATGGCTCGCAAGTTTAGGGAACAATCTTCTTCTTACCGTAGGAGG
AAGACGATGGTTTCTCTTAAGCCACTCATTAAACCTAATCCTCACATTTTCAATGATTTT
GACCTTACCCGTTAGCTTTGACCCACCTTGCTAGTAGTGACCATAAGCAGCTTATGTTA
CCAGTTCTTGTCTTCCAGGTTATGATCATGAAAATGAGAGTCCATTAAATGGTGGATATG
TTCGAAACCCAAATGATGGTTGGCGATTACATTGCATGGACACAAGAGGCAACTACATTC
GATTTCTTAAACCAACCGGGAAGAGTGAGATATTTGAAAGAATCAATGAGGAGAAGAAA
CCACCATTTTTCGATTTTCTTGGGTTGGGGACGGTGTGA

>G2424 Amino Acid Sequence (conserved domain in AA coordinates: 107-219)

MRMEMVHADVASLSITPCFPSSLSSSSHHYNNQQQHCHMSDQHHMSDQTTSSDYFSLNI
DNAQHLRSYYTSHREEDMNPNLSDYSNCKDITTVYRSCGHSSKASVSRGHWRAEDTKL
KELVAVYGPQNWNLIAEKLGSRGSKSRLRWFNQLDPRINRRAFTEEEERLMQAHRLYG
NKMAMIARLPFGRDTSVKNHWHVIMARKFREQSSSYRRRKTMSVSLKPLINPNPHIFNDF
DPTRLALTHLASSDHKQLMLPVPCFPGYDHENESPLMVDMFETQMMVGDYIAWTQEATTF
DFLNQTKKSEIFERINEEKKPPFFDFLGLGTV*

>G2505 (1..1026)

ATGGGTTCTTCGTCGAACGGAGGAGTGCCACCTGGTTTCCGGTTTTCATCCGACGGACGAA
GAGCTTCTCCTACTACTTGAAGAAGAAAATCTCTTACCAAAAGTTTGAGATGGAAGTC
ATCAGAGAGGTTGACTTAAACAAGCTTGAGCCTTGGGATTGCAAGAGAGATGCAAGATA
GGATCAACACCACAACGAATGGTACTTCTTCAGCCACAAGGACAGGAAATATCCGACG
GGGTCAAGGACCAACCGTGCTACTCATGCAGGTTCTGGAAGGCGACGGGACGTGACAAG
TGCAATAGGAACCTTACAAAAGATAGGAATGAGGAAGACACTTGTGTTCTACAAAGGT
AGAGCTCCTCATGGCCAAAAGACTGATTGGATCATGCATGAGTACCGTCTTGAAGACGCT
GATGATCCTCAAGCCAACCTAGTGAAGATGGATGGGTGGTATGTAGAGTGTTTATGAAG
AAAAATTTGTCAAGGTAGTAAATGAAGGTAGCTCAAGCATTAACCTCATTGGACCAACAC
AACCATGACGCATCTAACAACAACCATGCACCTCAAGCTCGTAGCTTTATGCACCGAGAC
AGTCCATACCAGCTAGTACGTAACACGGAGCCATGACATTGCAACTTAACAAGCCTGAC
CTTGCTCTTCATCAATACCCACCAATCTTCCACAAGCCACCTTCATTGGATTTGACTAC
TCTTCAGGACTTGCAAGGGACAGTGAGAGTGCGGCTAGTGAAGGGTTACAATACCAGCAA
GCGTGTGAGCCGGGTTTAGACGTTGGTACATGTGAGACAGTGGCTAGTCATAATCATCAA
CAAGGCTCAGGTGAATGGGCAATGATGGATAGACTTGTGACTTGTACATGGGAAATGAA
GATTCCTCTAGAGGGGATTACGTATGAGGATGGTAACAACAATTGCTCCTCTGTGGTTTCAG
CCAGTTCCTCGGACGAACAGCTAACATTGCGTAGTGAGATGGATTTCTGGGGTTATTCT
AAATAG

>G2505 Amino Acid Sequence (domain in AA coordinates: 10-159)

MGSSSNGGVPPGFRFHPTDELLHYLLKKISYQKFEMEVIREVLDNLKLEPWLQERCKI

TCTTCACCTCAATATTAGTAGCATGGAGGAGACTGTCTTATACTTTTTCGTTGTTAAGAT
AGGATTGGAGTGTCACTTAAGCTTGGAGGAGCTAACTCTGAAGTTCAGAAAAGCTTTGT
GTCTGATGAAGTGATCGTCTCTACCAATTAAAAACAAAATTCTACATGTACTAGAGCGTG
TATCGTTTTTTGGGATTAATAATCATATAATCGTTACATGAGCCTTGATACTTTGCTAGA
AATAAGCTCCTCTAAACAAAACCTTCTTTTTAAAAAACACACTTATGTTTTACTTAGTT
TGTTGTTGTATCCGAAGTTGATCAACGTTGTAATTTCCCAATAAATCATGACATTTTA
TATGCTCT

>G2519 Amino Acid Sequence (domain in AA coordinates:1-65)
MSHIAVERNRRRQMNEHLKSLRSLTPCFYIKRGDQASIIGGVIEFIKELQQLVQVLESKK
RRKTLNRPSFPYDHTIEPSSLGAATTRVPSRIENVMTTSTFKEVGACCNSPHANVEAK
ISGSNVVLRVVSRRIVGQLVKIISVLEKLSFQVLHLNISSMEETVLYFFVVKIGLECHLS
LEELTLEVQKSFVSDEVIVSTN*

>G2520 (133..1197)

AAGGAGTTTTGCATACTCACCAAGCCACAATCATTTCTCTCTTCTCTATCTCTCTGGTTT
TGAATCGGCGACGACTGAGTCAACTCGGTGTTGTTACTGGTTTCGTCGTATGTGTTGTAA
CTGATTAAGTTGATGGATCCGAGTGGGATGATGAACGAAGAGGACCGTTAATCTAGCG
GAGACTGGCAGTTTCCGTTGAACGGAGTTTCAACCGCCGAGATTCTTCTAGAAGAAGC
TTCGTTGGACCGAATCAGTTCGGTGATGCTGATCTAACCACAGCTGCTAACGGTGATCCA
GCGCGTATGAGTCACGCGTTGTCTCAGGCGGTTATTGAAGGTATCTCCGGCGCTTGAAAA
CGGAGGGAAGATGAGTCTAAGTCGGCGAAGATCGTCTCCACCATTGGCGCTAGTGAAGGT
GAGAACAAGACAGAGAATAGATGAAGTGTGTGATGGGAAAGCAGAAGCAGAATCGCTA
GGAACAGAGACGGAACAAAGAAGCAACAGATGGAACCAACGAAAGATTATATTCATGTT
CGAGCTAGAAGAGTCAAGCTACTGATAGTCACAGTTTGTAGTGAAGAGCGAGAAGAGAG
AAAATAAGTGAGCGGATGAAAATCTTGCAAGATCTTGTTCCGGGATGTAACAAGGTTATT
GGAAAAGCACTTGTCTAGATGAGATAATTAATATATACAATCATTGCAACGTCAAGTT
GAGTTCTTATCGATGAAGCTTGAAGCAGTCAACTCAAGAATGAACCTGGTATCGAGGTT
TTCCACCCAAAGAGGTGATGATTCTCATGATCATCAACTCAATCTTCTCCATTTTTTTC
ACAAAACAATACATGTTTCTATCGAGGTATTCTCGGGGTAGGAGTCTCGATGTTTATGCG
GTTCCGGTCATTTAAGCATTGCAATAAACGGAGTGACCTCTGTTTTGTCTCTGCTCCCCA
AAAACAGAACTTAAGACAACTATATTTTCAAAAACATGACATGTTTTCTGTCGATATTCT
CGAGTAGGAGTCCGCTATTAGTTTCATCTAAGCATTGCAATGAACCGTTTGGTCAGCAAGCG
TTTGAGAATCCGGAGATACAGTTCGGGTCGCAGTCTACGAGGGAATACAGTAGAGGAGCA
TCACCAGAGTGGTTCACATGCAGATAGGATCAGGTGGTTTCGAAAGAACGTCTTGA

>G2520 Amino Acid Sequence (domain in AA coordinates: 135-206)

MDPSGMMNEGPFNLAEIWQFPLNGVSTAGDSSRRSFVGPNQFGDADLTAAANGDPARMS
HALSQAVIEGIGSAWKREDESKSAKIVSTIGASEGENKRQKIDEVCDGKAEESLGTET
EQKKQMEPTKDYIHVRARRQATDSHSLAERARREKISERMKILQDLVPGCNKIVIGAL
VLDEIINYIQLRQVEFLSMKLEAVNSRMNPGIEVFPPEVMILMIINSIFSIFFTKQY
MFLSRYSRGRSLDVYAVRSFKHCNKRSDLCFCSCSPKTELKTTIFSQNMTCFCRYSRVGV
AISSSKHCNEPFGQAQAFENPEIQFGSQSTREYSRGASPEWLHMQIGSGGFERTS*

>G2533 (1..1080)

ATGATAAGCAAGGATCCAATATCGAGTTTACCTCCAGGGTTTCGATTTTCATCCAACAGAT
GAAGAACTCATTCTCCATTACCTAAGGAAGAAAGTTTCTCTTCCCCAGTCCCGCTTTTCG
ATTATCGCCGATGTCGATATCTACAAATCCGATCCATGGGATTACAGCTAAGGCTCCA
TTTGGGGAGAAAAGAGTGGTATTTTTTTCAGTCCGAGGGATAGGAAATATCCAACCGAGCA
AGACCAAACAGAGCAGCTGCGTCTGGATATTGGAAAGCAACCGGAACAGATAAATTGATT
GCGGTACCAAATGGTGAAGGGTTTCATGAAAACATTGGTATAAAAAAGCTCTTGTGTTT
TATAGAGGAAAGCCTCCAAAAGGTGTTAAAACCAATTGGATCATGCATGAATATCGTCTT
GCCGATTCTATCTCCAAAAGAATTAATCTTCTAGGAGCGGTGGTAGCGAAGTTAAT
AATAATTTTGAGATAGGAATTCTAAGAATATTGATGAGACTGGATGATTGGGTTCTT
TGCCGGATTTCAGAAATATCACAGCTTCATTGTCTATCACCTGATGTTGCTTTGGTCACA
AGCAATCAAGAGCATGAGGAAAATGACAACGAACCAATTCGTAGACCGCGGAACCTTTTG
CCAAATTTGCAAAATGATCAACCCCTTAAACGCCAGAAGTCTTCTTGTTCGTTCTCAAAC
TACTAGACGCTACAGATTGACGTTTCTCGCAAATTTTCTAAACGAAACCCGGAAAAT
CGTTCTGAATCAGATTTTTCTTTTCATGATTGGCAATTTCTCTAATCTGACATTTACGGA
AACCATTACTTGGATCAGAAGTTACCGCAGTTGAGCTCTCCCACTTCAGAGACAAGCGGC
ATCGGAAGCAAAAGAGAGAGAGTGGATTTTGCGGAAGAAACGATAAACCGCTTCGAAGAAG

ATGATGAACACATATAGTTACAATAATAGTATAGATCAAATGGATCATAGTATGATGCAA
CAACCTAGTTTCTGTAACAGGAACCTCATGATGAGTTCTCACCTTCAATATCAAGGCTAG
>G2533 Amino Acid Sequence (conserved domain in AA coordinates:11-186)
MISKDPISSLPPGFRFHPTEELILHYLRKKVSSSPVPLSIIADVDIYKSDPWDLPAKAP
FGEKEWYFFSPDRDKYPNGARPNRAAASGYWKATGTDKLIAPNGEGFHENIGIKKALVF
YRGKPPKGVKTNWIMHEYRLADSLSPKRINSSRSGGSEVNNNFGRNSKEYSMRLDDDWVL
CRIYKKSHASLSSPDVALVTSNQHEENDNEPFDVDRGTFLPNLQNDQPLKRQKSSCSFSN
LLDATDLTFLANFLNETPENRSEDFSMIGNFSNPDIYGNHYLDQKLPQLSSPTSETSG
IGSKRERVDAEETINASKMMNTYSYNNSIDQMDHSMMQQPSFLNQELMMSSHLQYQG*
>G2534 (1..975)
ATGGATAATATAATGCAATCGTCAATGCCACCGGGATTCCGATTTCATCCGACAGAGGAA
GAGCTTGTGGGTATTACCTAGATAGGAAGATCAATTCATGAAGAGTGCTTTAGATGTC
ATTGTAGAGATTGATCTCTACAAAATGGAGCCATGGGATATACAAGCGAGGTGTAACCTA
GGGTATGAAGAGCAAAACGAGTGGTACTTCTTTAGTCATAAGGACAGGAAGTACCCTACC
GGGACTAGGACCAACCGAGCCACTGCGGCTGGGTCTTGAAAGCCACGGGTAGAGACAAG
GCGGTACTATCAAAAAACAGTGTCTATCGGAATGCGGAAGACACTTGTCTACTACAAGGT
CGAGTCTCTAATGAAGAAAGTCCGATTGGATCATGCACGAATACCGTCTCCAAAACCTCC
GAGCTTGCCCCGGTTTCAGGAGGAAGGCTGGGTGGTGTGTCGAGCATTAGGAAGCCAATT
CCAAACCAGAGGCCATTAGGGTACGAGCCATGGCAGAACCAGCTCTACCACGTGAAAGT
AGTAACAACCTACTCATCTTCAGTGACAATGAACACGAGTCATCATATCGGTGCATCTTCA
TCAAGTCATAACCTTAATCAAATGCTCATGAGCAATAACCACTACAATCCTAATAATACA
TCCTCATCGATGCATCAATATGGCAACATTGAGCTCCCGCAGTTGGACAGCCCGAGCTTG
TCGCCTAGTTTGGGACGAATAAAGATCAGAACGAGAGTTTCGAGCAAGAAGAAGAGAAG
AGCTTTAAGTGTGTGGATTGGAGAACACTAGATACCTTGCTTGAGACACAAGTCATACAT
CCGCATAACCCTAATATTCTTATGTTGAAACGCGAGTCGTATAATCCGGCGCCAAGCTTC
CCTTCCATGCATCAAAGCTATAATGAGGTCGAAGCTAATATTCATCATTTCTTGGATGC
TTCCCTGACTCGTAA
>G2534 Amino Acid Sequence (conserved domain in AA coordinates:10-157)
MDNIMQSSMPGFRFHPTEELVGYLDRKINSMSALDVIVEIDLYKMEPWDIQARCKL
GYEEQNEWYFFSHKDRKPYTGTRTNRATAAGFWKATGRDKAVLSKNSVIGMRKTLVYK
RAPNGRKSQWIMHEYRLQNSLAPVQEEGWVVCRAFRKPIPNRPLGYEPWQNLVHVES
SNYSSSVTMTSHHIGASSSSSHNLNQLMSNNHYNPNNTSSSMHQYGNIELPQLDSPSL
SPSLGTNKDQNESFEQEEKSFNCVDWRTLDTLLETQVIHPNPNILMFETQSYNPAPSF
PSMHQSYNEVEANIHHSLGCFPDS*
>G2573 (34..957)
CCAGATTAAATTTGAGACTCTCAAAGAAACACCATGGAAGAAGAGCAACCTCCGGCCAAG
AAACGAAACATGGGGAGATCTAGAAAAGGTTGCATGAAAGGTAAAGCGGTCAGAGAAC
GCCACGTGTACTTTCCGTGGAGTTAGGCAACGGACTTGGGGTAAATGGGTGGCTGAGATC
CGTGAGCCTAACCGTGGGACTCGTCTCTGGCTCGGCACGTTTAAATACCTCGGTGAGGCC
GCCATGGCTTACGATGAAGCCGCTAAGAAACTCTATGGACACGAGGCTAAACTCAACTTG
GTGCACCCACAACAACAACAAGTAGTAGTGAACAGAAACTTGTCTTTTTCTGGCCAC
GGGTGCGGTTCTTGGGCTTATAATAAGAAGCTCGATATGGTTCATGGGTGGACCTTGGT
CTCGGCCAGGCAAGTTGTTACGAGGTTCTTGCTCAGAGAGATCGAGTTTCTACAAGAA
GATGATGATCATAGTCATAATCGATGTTCTGCTCTCAAGTGGTTCGAATCTTTGTGGTTA
TTACCTAAACAAAGTGATTACAAAGATCAAGAGACCGTTAATGCTACGACTAGTTATGGC
GGTGAAGGCGGTGGTGGCTCTACGTTAACGTTTTCGACCAATTGAAACCAAGAATTTG
ATGAGTCAGAATTATGGATTATACAATGGAGCTTGGTCTAGGTTCTTGTGGGGCAAGAA
AAGAAGACGGAACATGACGTGTCTCGTCTGTTGGATCGTTCGACAACAAGGAGAGTATG
TTGGTTCCTAGTTTCGGCGGAGAGAGGATGCATAGGCCGGAGTTGGAAGAGCGAACAGGA
TATTTGGAATGGATGATCTTTTGGAGATTGATGATTAGGTTTGTGATTGGCAAAAAT
GGAGATTTCAAGAATTGGTGTGTGAAGAGTTTCAACATCCATGGAATTGGTTCTGAGAG
TTTTTATTTATTACTATTATTATTCATACATATTTCTTATATTGACTTAGG
>G2573 Amino Acid Sequence (domain in AA coordinates: TBD)
MEEQPPAKKRNRMRGRKGMKGKGGPENATCTFRGVRQRTWGKQWAEIREPNRGRRLWL
GTFNTSVEAAMAYDEAAKLYGHEAKLNLVHPQQQQQVVNRNLSFSGHSGSGSWAYNKKL
DMVHGLDLGLGQASCSRSERSSFLQEDDDHSHNRCSSSSGSNLCWLLPKQSDSQDQE
TVNATTSYGGEGGGSTLTFSTNLKPKNLMSQNYGLYNGAWSRFLVGQEKTEHDVSSSC

GSSDNKESMLVPSCGGERMHRPELEERTGYLEMDDLLEIDDLGLLIGKNGDFKNWCCEEF
QHPWNWF*

>G2589 (23..1354)

AAAGAAAAGAAAAATAAGATAATGAGGACGAAGACTAAGTTAGTACTCATACCTGATAG
ACACTTTCGGAGAGCCACATTCAGGAAGAGGAATGCAGGGATAAGGAAGAACTCCACGA
GCTGACAACTCTCTGTGACATCAAAGCATGTGCGGTAATCTACAGTCCGTTTCGAGAATCC
AACGGTGTGGCCGTCAACCGAAGGTGTCAAGAGGTGATTCGGAGTTCATGGAGAAGCC
GGCGACAGAACGGTCCAAGACGATGATGAGTCATGAGACTTCTTTCGCGGACCAAATCAC
CAAAGAACAAAACAACTAGAGAGTCTACGTCGTGAAAACCGAGAACTCAGCTTAAGCA
TTTTATGTTTGATTGCGTTGGAGGCAAGATGAGTGAGCAACAGTATGGTGCAAGGGACCT
TCAAGATTAAAGTCTTTTACTGATCAATATCTTAATCAGCTTAATGCCAGGAAGAAGTT
CCTTACAGAAATATGGTGAGTCTTCTTCTTCTGTTCTCTCTGTTTGTATGTTGCGGGTGC
CAATCCTCCTGTTGTTGCAGATCAAGCTGCGGTAAGTGTCTCTCTTGTGTTGCTGTTGC
GGGTGCCAATCTTCTGTTGTTGCTGATCAAGCTGCGGTAAGTGTCTCTCTGTTTGC
TGTTGCGGGTGCCAATCTTCTGTTGTTGTCAGATCAAGCTGCGGTTAATGTTCTCTACTGG
ATTTTCATAACATGAATGTGAACAGAAATCAGTATGAGCCGGTTCAGCCCTATGTCCCTAC
TGTTTTTAGTGATCATATCAATATCAGAAATATGAAGTCAATCAAAACCAACAAGAGCC
GGTTCATTACCAAGCTCTTGTCTGTTGCGGGTCCGGTCTTCTATGACTCAGAATCAGTA
TGAGCCCGTTCACTACCAGAGTCTTGTGTCGCGGGTGGCGGTCTTCTATGAGTCAGTT
GCAGTATGAGCCGGTTCAGCCTTATATCCCTACTGTTTTTAGTGATAATGTTCAATATCA
GCATATGAATTTGTATCAAAATCAACAAGAGCCGGTTCCTACCAAGCTCTTGGTGTGTC
AGGTGCCGGTCTTCTCTATGAATCAGAAATCAGTATGAGCCGGTTCAGCCCTATGTCCCTAC
TGTTTTTAGTGATCATTTTCAGTTTGAGAATATGAATTTGAATCAAAATCAACAGGAGCC
GGTTCAATACCAAGCTCTGTTGATTTTAATCATCAGATTCAACAAGGAACTATGATAT
GAATTTGAACAGAAATGAGTTTGGATCCAAATCAGTATCCGTTTCAAAATGATCCATT
CATGAATATGTTGACAGAAATATCCTTATGAATAAGCGGGTATGTTGAGAGCATGCAC

>G2589 Amino Acid Sequence (domain in AA coordinates: TBD)

MRTKTKLVLPDRHFRRAFRKRNAGIRKKLHELTTLCDIKACAVIYSPFENPTVWPSTE
GVQEVISEFMEKPATERSKTMMSHETFLRDQITKEQNKLES LRRENRETQLKHFMFDCVG
GKMSEQQYGARDLQDLSLFTDQYLNQLNARKKFLTEYGESSSSVPLFDVAGANPPVAD
QAAVTVPPLFAVAGANLPVVADQAAVTVPPLFAVAGANLPVVADQAAVNVPTGFHNMNVN
QNQYEPVQPYVPTGFSDFHLYQYQNMNFQYQEPVHYQALAVAGAGLPMTQNQYEPVHYQS
LAVAGGGLPMSQLQYEPVQPYIPTVFSQYQYQHMNLYQYQEPVHYQALGVAGAGLPMN
QNQYEPVQPYVPTGFSDFHLYQYQNMNLYQYQEPVQYQAPVDFNHQIQQGNVDMNLYQNMMS
LDPNQYPFQNDPFMNMLETPYE*

>G2687 (45..1139)

CTCTGTCTCTCGTATCTTTCTACTACTCTGTTTCTTGAATTCTAATGAACAACATCGACG
ACGCAAGACGGAGACTTCAGTGTCTTCAGGTTCAAGCGACTCTTCTTGCCTCTCAAGA
AACGCGATGAGACTTGATGACGAACCAAGCAACCGCCTAGTGTTTCTGCTTTCACCAAGA
CGGTTGTGGCTTCTGGCAATGTCAAGTACAAAGGAGTCGTTTCAGCAACAGAACGGTCATT
GGGGTGCCAGATTTACGCAGACCACAAAAGGATTTGGCTTGGAACCTTCAAATCCGCTG
ATGAAGCCGCCACGGCTTACGATAGTGCATCTATCAAACCTCCGAAGCTTTGACGCTAACT
CGCACCAGAACTTCCCTTGGTCTACAACTCACTCTCAACGAACAGACTTTCAAATTTGCT
ACACAACAGAGACTGTGTTGAACATGATCAGAGACGGTTTCGTACCAACACAAATTCAGAG
ATTTTCTCAGAATCAGATCTCAGATTGTTGCGAGTATCAACATCGGGGGACCAAAACAAG
CCCGAGGAGAAGTGAATCAAGAATCAGACAAGTGTTTTCTTGACACAGCTTTTTCAGA
AGGAATTGACACCGAGCGATGTAGGGAACTAAATAGGCTTGTGATACCTAAAAAGTATG
CAGTGAAGTATATGCTTTTCATAAGCGCTGATCAAAGCGAGAAAGAAGAGGGTGAAATAG
TAGGATCTGTGGAAGATGTGGAGGTTGTGTTTTACGACAGAGCAATGAGACAATGGAAGT
TTAGGTATTGTTACTGGAAGAGTAGCCAGAGCTTTGTCTTACCAGAGGATGGAATAGTT
TCGTGAAGGAGAAGAATCTCAAGGAGAAGGATGTTATTGCCTTCTACACTTGCAGATGTC
CGAACAAATGTGAAGCATTAGAAGGTCAAAGAAAGAACTTCTTGATGATCGATGTTTCATT
GCTTTTTCAGACAACGGTTCCGTTGAGCTGAGGAAGTAAGTATGACGGTTCATGACAGTT
CAGTGCAAGTAAAGAAAACAGAAAACCTTGGTTAGCTCCATGTTAGAAAGATAAAGAAACCA
AATCAGAGGAGAACAAGGAGGGTTTATGCTGTTTGGTGTAAGGATCGAATGTCCTTAGG
GAATTTTCTTTAAAGTTTCTTACTTCAACTAGAACTTGTTTTACTTGTTACCT

>G2687 Amino Acid Sequence (domain in AA coordinates: TBD)

MNNIDDAKTETSVSSGSSDSFLPLKKRMRLDDEPENALVVSSSPKTVVASGNVYKGVVQ
QQNGHWGAQIYADHKRIWLGTFKSADAEATAYDSASIKLRSFDANSHRNFPWSTITLNEP
DFQNCYTTETVLNMIRDGSYQHKFRDFLRIRSQIVASINIGGPKQARGEVNQESDKCFSC
TQLFQKELTPSDVGKLNRLVIPKKYAVKYMPPISADQSEKEEGEIVGSVEDVEVVYFDRA
MRQWKFRYCYWKSSQSFVFTRGWNSFVKEKNLKEKDVIAYFTCDVPPNNVKTLEGQRKNFL
MIDVHCFSDNGSVVAEEVSMTVHDSSVQVKKTENLVSSMLEDKETKSEENKGGFMLFGVR
IECP*

>G27 (83..622)

CAAAATACCAAAAAACAAACATTTTTTTTAACTCTCCCAACATTTTTTTCTCTTCTCT
CGTTACATTAAATTATCTTTAGATGCAAGACTCTTCTCTCACGAATCGCAACGTAACCT
CCGGTCACCGGTGCGCGAGAAAACCGGAAAGAGTTCTAAGACTAAAAATGAGCAAAAAGG
TGTTTCTAAACAACCAATTTTCGTGGGGTCAGAATGAGACAATGGGGAAAATGGGTGTC
TGAAATTAGAGAACCAAGAAAGAAATCAAGAATATGGCTCGGTACTTTCTCTACGCCGGA
GATGGCGGCGCGTGACACGACGTGGCGGCTTTAGCCATCAAAGGTGGCTCTGCCACCT
TAATTTCCCGGAGCTAGCTTACCATTGCGGAGACCGGTAGCGCGGACCCCTAAAGACAT
TCAAGAAGCCGCGCGCAGCAGCTGCCGTTGACTGGAAAGCACCGGAGTCTCCGTCTAG
CACCGTGACGTCTCTCCAGTCGCGGACGACGCTTTCTCCGATCTTCTGATCTTTTGCT
TGACGTGAATGATCACAACAAAACGATGGATTCTGGGACTCGTTTCCGTACGAAGATCC
TTTCTTCTTGAAAAATTACTAGAAGGCAAATCTTGCCGGCGAACGGATTTTCCGGTGGT
TTCCCGGTAAATAAGAAGACGATGTCGTTTTGTACCTTTTTTGTCTACGATGGGAAATTT
CTTTTTTTTTTACGTGTGAGTAAAGTTTCCGAATGTGTGATGTGTAAGTAAGTACAGGT
TATTTAATTTCTTTTTTTTTGTACAAATACGTACGTATTACCAAAAAGTTTTCATTTATT
GTGCTTTTATCTTCCAAATTCATTAAAAAAGTAAAAAAGTAAAAAAGTAAAAAAGT

>G27 Amino Acid Sequence (domain in AA coordinates: 37-104)

MQDSSSHESQRNLRSPVEKTKGSSKTKNEQKGVSKQPNFRGVRMRQWGKVVSEIREPRK
KSRIWLGTFSTPEMAARAHDVAALAIKGGSAHLNFPDELAYHLPRPASADPKDIEAAAAA
AAVDWKAPESPSTVSSPVADDAFSDLPDLLLLVDNDHNKNDGFWDSPFYEDPFFLENY*

>G2720 (1..894)

ATGGAAGCGAAGAAGGAAGAGATAAAGAAAGGTCCATGGAAAGCCGAAGAAGACGAAGTA
CTCATCAACCATGTCAAGAGATACGGTCCTCGTGATTGGAGCTCCATTTCGATCCAAAGGT
CTTCTTCAACGCACCGGCAAATCTGTCTCTCGTTGGGTCAATAAACTCCGTCCCAAT
CTCAAAAATGGATGCAAGTTCTCGCTGACGAAGAGAGGACTGTGATTGAGTTACAATCT
GAGTTTGGTAACAAATGGGCGAGAATCGCTACGTATCTACCGGGAAGAAGTGAACGAT
GTGAAGAATTTCTGGAGTAGCAGACAAAAGAGACTCGCTAGGATTCTTCATAACTCCTCT
GATGCATCGAGTTTCGAGTTTCAATCCCAATCTTCTTCTCTCATCGACTCAAGGGCAA
AACGTCAAACCAATCCGTCAATCTCTCAGGGTTTTGGTTTGGTTGAGGAAGAGGTTACA
GTTTCTTCTTCATGTTCCAGATGGTTTCTTATTCTATCTGATCAAGTTGGTGATGAAGTC
TTGAGGTTGCCGATTTGGGTGTTAAGTTAGAGCATCAGCCTTTTCGCTTTTGGCACTGAT
CTTGTCCTAGCAGAGTACTCTGACTCACAGAATGATGCAAATCAGCAAGCAATCAGCCCT
TTCTCTCAGAAAGCAGAGAGCTTTTGGCTAGACTTGACGACCCTTTTACTATGATATA
CTTGGACCGCTGATTCTTCTGAGCCATTGTTTCGCTCTCCCTCAGCCGTCTTTCGAGCCT
TCGCCTGTGCCGAGAAGATGCAGACATGTTTCAAAGGATGAAGAAGCTGATGTTTTCTTA
GACGATTTCCAGCTGACATGTTTGATCAGGTTGATCCAATCCCAAGTCCTTAG

>G2720 Amino Acid Sequence (domain in AA coordinates: 10-114)

MEAKKEEIKKGPWKAEEDEVLINHVKRYGPRDWSSIRSKGLLQRTGKSCRLRWVNLKLPN
LKNGCKFSADEERTVIELQSEFGNKWARIATYLPGRTDNDVKNFWSSRQKRLARILHNS
DASSSFFNPKSSSSHRLKGKNVKPIRQSSQGFGLVEEEVTVSSSCSQMVPYSSDQVDEV
LRLPDLGVKLEHQPFAGFDLVLAEYSDSQNDANQQAISPFSPESRELLARLDDPFYDI
LGPADSSEPLFALPQPFEPSPVPRRHRVSKDEEADVFLDDFPADMFDQVDPISP*

>G2787 (142..1584)

TCTCAGAGCAAAAAACAAAAAAGAAAAAACCCTAAATCTAAATCTCACCTTCCA
CCTCTGTCTTTTTTTTTTTGTTCTTTTTTTTTTTTACTGTATCTTCTCTCTTTG
CTCTGCAAAAATCTCACATCCATGGATCCATCTCTTGGTGATCCTCATCATCCTCCTCAG
TTCACCCCTTTTCTCATTTTCCCACCTCAAATCATCATCTTTAGGACCAAATCCGTAC
AATAACCATGTCGCTTCCAACCGCAGCCGCAACGCAACGCAAAATCCCGCAACCGCAG
ATGTTTCAGTTATCTTCCACATGTTTCAATGCCCAACCTCCTTACTCCGAAATGATTTGC
GCTGCGATTGCGGCGTTAAACGAACCGGATGGTTCGAGCAAGATGGCAATTTTCGAGATAC

ATCGAGAGATGTTACACCGGTTTAACTTCTGCTCATGCTGCTTTGTTGACTCACCATCTC
AAGACTTTGAAGACCAGTGGTGTCTTTCTATGGTTAAGAAATCTTACAAAATTGCTGGT
TCTTCTACTCCTCCTGCTAGTGTAGCTGTGCTGCTGCTGCCGCCGCTCAAGGTCTCGAT
GTTCCAGATCTGAGATTCTCCATTCAAGTAACAACGATCCCATGGCTTCTGGCTCTGCT
TCTCAGCCTCTGAAACGAGGTCGTGGTCGTCCTCTAAGCCTAAACCTGAATCTCAACCA
CAACCACTACAGCAACTTCCACCGACCAATCAAGTCCAGGCTAACGGACAGCCAATCTGG
GAACAGCAGCAAGTTCAATCACCTGTTCCGGTTCGACTCCGGTTACAGAGTCGGCGAAG
AGAGGACCTGGTTCGTCGAAGGAAGAACGGTCTGCTGCTCCTGCTACTGCACCAATCGTT
CAAGCTTCGGTTATGGCTGGAATTATGAAACGTAGAGGTAGACCACCGGGTCGTGAGCT
GCTGGGAGACAGAGGAAGCCCAAATCCGTTTCTTCTACTGCCTCTGTGTATCCTTATGTT
GCTAATGGTGTAGACGCAGAGGAAGGCCTAGGAGAGTTGTTGACCCTAGCAGTATTGTT
AGTGTGCTCCAGTAGGTGGTGAAATGTGGCAGCGGTTGCGCCAGGGATGAAGCGTGGA
CCTGGACGACCACCTAAGATTGGTGGTGTATCAGTAGGCTTATTATGAAGCCTAAGAGA
GGACGAGGACGTCCTGTAGGTAGACCCAGAAAGATTGGAACATCAGTCACGACTGGGACA
CAAGATTCTGGAGAACTCAAGAAGAAGTTTGATATTTTCAAGAGAAAGTGAAAGAAAT
GTGAAGGTGTTGAAGGTGAGTTACAAGTGAGAATCAAGCAGTGGTGCAAGCCATAAAA
GATCTGGAAGCACTAACAGTGACGGAGACCGTTGAGCCACAAGTTATGGAAGAAGTGCAG
CCAGAGGAGACTGCAGCACCACAGACTGAAGCTCAACAACTGAAGCTGCTGAGACACAA
GGAGGACAAGAAGAAGGACAAGAAAGAGAAGGAGAAACACAGACCCAGACAGAAGCAGAG
GCAATGCAAGAAGCTCTGTTCTGAAGAATAATAATGATCTAGAAAACAACCTAGACATAA
TAGCCTTGTTGTTGGCGTTAGGAGTGTCTTTTTTTAGTTGTTTGGTGTGGAATCGC
ATCTTAAATTATATAAAAATCTATAAGGAATTTTAAATTTTCTAGGTTTGTGCTGCA
GAAGAAGAAATAGTAGACTCGTTAATGGTGTGTTGTCGGTGTGCTTTAACCAAAACCAT
AAGACGTGGCTGTAAATTAGCGATGTTTCTAGTCTTCCATCTTTAATAATCTCTTATTGC
GTCTGTGCCTTTGTTTTT

>G2787 Amino Acid Sequence (domain in AA coordinates: 172-192, 226-247, 256-276, 290-311, 245-366)

MDPSLGDPHHPQFTFPFHPFPTSNHHPLGPNPYNHVVFPQPQTQTQIPQPQMFQLSPH
VSMPPHPYSEMICAIAALNEPDGSSKMAISRYIERCYTGLTSAHAALLTHHLKTLKTSG
VLSMVKKSYKIAGSSTPPASVAVAAAAAQGLDVPREILHSSNNDPMASGSASQPLKRG
RGRPPKPKPESQPQLQPLPTNQVQANGQPIWEQQVQSPVPVPTPVTESAKRGPGRPR
KNGSAAPATAPIVQASVMAGIMKRRRPPGRRRAAGRQRKPKSVSTASVYPYVANGARRR
GRPRRVDPSSIVSVAPVGGENVAAPVPMKRRGRPPKIGGVISRLIMKPKRGRGRPVG
RPRKIGTSVTTGTQDSGELKKKFDIFQEKVKEIVKVLKDGVTSENQAVVQAIKDLEALTV
TETVEPQVMEEVQPEETAAPQTEAQQTAAETQGGQEEGQEREGETQTQTEAEMQEALF
*

>G2789 (82..879)

CTTTAGGGACACCAAATCTATTCAACCTAAAAGCCTTCTTTTCCCTATATTGACCAACT
TTTTAGCGAATCAGAAGAGGAATGGATGAGGTATCTCGTTCTCATACACCGCAATTTCTA
TCAAGTGATCATCAGCACTATCACCATCAAAACGCTGGACGACAAAAACGGGCAGAGAA
GAAGAAGGAGTTGAACCCAACAATATAGGGGAAGACCTAGCCACCTTTCTTCCGGAGAA
GAGAATATCAAGAAGAGAAGGCCACGTGGCAGACCTGCTGTTCCAAGAACAACCCAAA
GCACCAATCATAGTCACTCGCGACTCCCGGAACGCTTCAGATGTACGTCATGGAGATA
ACCAACGCTGCGATGTAATGGAAGCCTAGCCGTCTTCGCTAGACGCCGTGAGCGTGGC
GTTTGCCTCTTGACCGGAAACGGGGCGTTACAAACGTCACCGTTAGACAACCTGGCGGA
GGCGTCGTCAGTTTACACGGACGGTTTGAGATTCTTCTCTCTCGGGTTCTGTTTCTCTCT
CCACCGGCACACCAAGCTGCGTCTGGTTTAAAGGTTTACTTAGCCGGTGGTCAAGGTCAA
GTGATCGGAGGCAGTGTGGTGGGACCGCTTACGGCATCAAGTCCGGTGGTCTGTTATGGCA
GCTTCATTTGGAACGCATCTTACGAGAGGCTGCCACTAGAGGAGGAGGAGAACTGAA
AGAGAAATAGATGGAACCGCGCTAGGGCGATTGGAACGCAACGCAGAAACAGTTAATG
CAAGATGCGACATCGTTTATTGGGTCGCCGTCGAATTTAATTAACCTCTGTTTCGTTGCCA
GGTGAAGCTTATTGGGGAAACGACACCGCTCTTCTAAGATAATATCATTGATAATATA
AGTTTCGTCTTCTTATCTTTTTTCACTTTTTTACCTTTTTTCACTTCTTAGGTTTGTGTTT
AACGTTTGATTAATACCTGAAGGTTTTTGGAAAATTTTCGATCGGATAAAAGGATTATG
TTGCGAGCCGAAACGCGGCC

>G2789 Amino Acid Sequence (domain in AA coordinates: 53-73, 121-165)
MDEVSRSHTPQFLSSDHQHYHHQNAQRQKRGREEGVENNIGEDLATFPSGEBENIKRR

PRGRPAGSKNPKAPIIVTRDSANAFRCHVMEITNACDVMESLAVFARRRQRGVCVLTGN
GAVTNVTVRQPGGGVVS LHRFEILSLSGSFLPPPAPPAASGLKVYLAGGQGVIGGSV
GPLTASSPVVMAASFGNASYERLPLEEEEEETEREIDGNAARAIGTQTQKQLMQDATSFI
GSPSNLINSVSLPGEAYWGTQRPSF*

>G31 (13..615)

CTTTTATAAGCAATGGCTCCAAGACAGGCGAACGGTAGAAGCATTGCCGTGAGTGAAGGC
GGCGGAGGGAAGACGATGACGATGACGACGATGCGGAAGGAAGTGCACCTTTAGAGGTGTG
AGGAAGCGTCCATGGGGTAGATACGCGGCGGAGATCCGTGACCCGGGAAAGAAAACCCGG
GTTTGGCTCGGGACATTTCGACACGGCGGAGGAAGCTGCAAGAGCTTACGACACCGCCGCT
AGAGAGTTTCGTGGCTCCAAAGCAAAGACTAATTTCCCTCTTCCCGGAGAGTCTACTACG
GTTAACGACGGTGGCGAGAACGATTCTTACGTCAACCGTACGACGGTGACGACGGCGCGT
GAGATGACGCGTCAGAGATTTCCGTTTGCATGTACCCGGGAGCGTAAAGTCGTCGGTGGT
TATGCTTCTGCTGGTTTTTCTTCGATCCGTCAAGAGCTGCTTCGTTAAGAGCAGAGCTT
TCTCGGGTTTGTCCGGTTCCGGTTTGTATCCGGTTAATATCGAGTTGAGTATTGGTATTCTGA
GAAACCGTAAAAGTTGAACCGAGAAGAGAATAAACCCTGGATCTTAACCTAGCTCCACCG.
GTGGTGGACGTTTAGATTTTTTCTTCTTTTCATAATTTGTATTTTACATTGCCGGA
TAATTAATGTTTTCTTTAG

>G31 Amino Acid Sequence (domain in AA coordinates: TBD)

MAPRQANGRSIAVSEGGGGKMTMTTMRKEVHFRGVKRPWGRYAAEIRDPGKKTRVWL
TFDTAEAAARAYDTAAREFRGSKAKTNFPLPGESTTVNDGGENDSYVNRITVTAREMTR
QRFPFACHRRERKVVGGYASAGFFDPSRAASLRAELSRVCPVRFPVNIELSIGIRETVK
VEPRRELNLDLNLAPPVVDV*

>G33 (20..757)

ATTCTCCCCCAACCAAAATATGACCACAGAAAAAGAGAATGTCACTACGGCCGTGGCCGT
GAAAGACGGCGGAGAAAAGAGTAAGGAAGTGAGTGACAAGGGCGTAAAGAAGAGAAAGAA
TGTAACCTAAGGCCCTGGCCGTGAATGACGCGGAGAAAAGAGTAAGGAAGTGCCTTACAG
GGGTGTAAGGAGGAGACCATGGGGGAGATATGCTGCGGAGATCCGTGATCCCGTAAAGAA
AAAACGGGTCTGGCTCGGGTCTTCAACACGGGGGAGGAAGCCGCCAGAGCCTACGACTC
CGCTGCCATAAGGTTTCGAGGATCGAAAGCTACTACTAATCTCCCTCTAATCGGATACTA
TGCGGATTTCTTCGCGCAGCGCCGTGAACAACAACCTTTCCGAGACGGTGAGTGATGGAAA
TGCCAACCTCCCTCTCGTTGGAGACGATGGGAATGCTTTGGCTTCTCCGGTGAACAACAC
CCTTTCCGAAACGGCGCGTGATGGAAACACTTCCATCGGATTGTACGACATGTTATCTCC
GGGGGTGGCTGAAGCGGTTGCTGGATTTTTCTTAGATCTGCCTGAAGTTATTGCGTTGAA
AGAGGAGCTTGATCGAGTTTGTCTGACAGTTTGTAGTCCATTGATATGGGGTTGACTAT
TGGTCTCTCAAACCGCCGTGGAGAGCCTGAGACTTCTCCGCCGTGGATTGTAAGCTGCG
AATGGAACCGGATCTTGACCTCAACGCAAGTCCCTAAAGATTGATCTGATGTTGTTGTCC
TTGAATAAGTTTGTATCTTGTGCTCTTCTGATTGTCTGTACTTCTATTGGTTGATTGCG
TGCTTTTGGAGGACAAAACAAACATTTTTTTATGTATTAAAAAAGGTAATTGAACTATT
ATCGAAAAA

>G33 Amino Acid Sequence (domain in AA coordinates: 50-117)

MTEKENVTAVAVKDGGEKSKEVSDKGVKRKNVTALAVNDGGEKSKEVRYRGVRRRP
WGRYAAEIRDVPVKKRVWLGSFNTGEEAARAYDSAAIRFRGSKATTNPLIGYYGISSAT
PVNNNLSETVSDGNANLPLVGDDGNALASPVNNTLSETARDGTLPSDCHDMLSPGVAEAV
AGFFLDLPEVIALKEELDRVCPDQFESIDMGLTIGPQTAVEEPETSSAVDCKLRMEPDLD
LNASP*

>G342 (1..723)

ATGGACGTCTACGGCATGTCTTCACCGGACTTGCTTCGTATCGACGACCTTCTCGATTTC
TCCAACGACGAAATTTCTCTCTTCCACCGTCACTTCTCCGCCGCTTCTCTCCGCC
GCTTCTTCCGAAAACCTTTCAGCTTTCTCTTCCACCTACACTTCTCTACTCTCTCTC
ACCGACTTCACTCACGATCTCTGCGTTCCAGTGACGACGAGCTCATCTCGAATGGTTA
TCGCGATTGCTTGACGATTCACTTCCGATTTCAGCAAAATCTTTAACCATGACCGTT
AGACCCGAGATTTCACTACCCGAAAACCTAGAAGTCGCCGATCAAGAGCACCAGCACCT
TCCGTAGCTGGAACCTGGGCTCCGATGTCTGAATCAGAGCTTTGTCACTCCGTCGCTAAA
CCTAAACCGAAGAAAGTCTACAACGCTGAATCGGTTACGGCGGATGAGCGGAGCGGTGC
ACGCACTGTGCCTCGGAGAAAACGCCACAGTGGAGAACTGGACCGCTTGGACCTAAAACA
CTTTGTAAACGCTTGTGGAGTTCTGTTACAAATCAGGGAGGCTTGTACCGGAATACAGACCG
GCGTCGAGTCCGACGTTTGTATTGACTCAGCACTCGAACTCTCATCGGAAAGTTATGGAG

>G342 Amino Acid Sequence (domain in AA coordinates: 155-190)
MDVYGMSSPDLLRIDDLLDFSNDIEIFSSSSTVTSSAASSAASSENPFSPSTSTYTSPTLL
TDFTHDLCVPSDDAHLEWLSRFVDDSFSDFPANPLTMTVRPEISFTGKPRSRRSRAPAP
SVAGTWAPMSESELCHSVAKPKPKKVYNAESVTADGARRCTHCASEKTPQWRGTGLGPKT
LCNACGVRYKSGRLVPEYRPASSPTFVLTHQSHNSHRKVMELRRQKEQQESCVRIPPFQPQ

[illegible]

MALETLSNPSTATTARPLLRYREEMEPENLEQWAKRKRTKRQRFDHGHQNETKNLPSE
EYYLALCLMLLARGSAVQSPPLPPLPSRASPSDRDYKCTVCCKGSFSSYQALGGHKTSHR
KPTNTSITSGNQELSNNSHNSGVSVINVTNTNGVVSQSGKIHTCSICFKSFASGQALG
GHKRCHYDGGNNNGNGSSSNSVELVAGSDVSDVDNERWSEESAIGGHRGFDLNLPAQV
SVTTS*

ATGCAGAACAAACACAAATGCAAGCTCTGTTCCAAGAGTTTCTGTAATGGCAGAGCACTT
GGTGGTCCACATGAAGTCTCACTTGGTCTCATCTCAGTCTTCAGCTCGGAAGAACTAGGT
GACTCGGTCTATTCTTCTTCTTCTTCTTCTCCGATGGTAAAGCGCTCGCCTACGGGTTA
CGAGAGAACCCGAGGAAGAGTTTCCGGGTCTTTAATCCGGATCCTGAGTCATCCACAATT
TACAACAGTGAGACAGAGACCGAACCTGAATCCGGAGACCCGGTTAAGAAACGGGTGAGA
GGAGATGTTTCAAAGAGAAGAAGAGAAGGCAAAGAGTAAGAGAGTGTGTTGAGAACTCG
AAGAAGCAAAGACAATTACGAGTCAACCAGAACCGAGGTTCTGTCTCTGATGGTTCT
CCTGAACAAGATTTAGCTATGTGCTTGATGATGCTGTCAAGAGATTCAAGGGAGCTCGAG
ATTAACCTGAAAAAACCCGAGGAAGAGAGGAAGCCGGAAAAAGACATTTCCCTGAGCTC
CGTCTGCTATGATAGATCTGAATCTTCTCCGCCGCAAGAAGCTGAAGCTGTCAACCGTC
TTTTCAGCTATATAA

MQNKHKCKLCSKSFNCGRALGGHMKSHLVSSQSSARKKLGDVYSSSSSSSDGKALAYGL
RENPRKSFVRVFPDPESPSTIYNSETETEPESGDPVKKRVRGDVSKKKKKKAKSKRVFENS
KKQKTIHESPEPASSVSDGSPEQDLAMCLMMLSRDSRELEIKLKKPEERKPEKRHPPEL
RRCMIDLNLPPPOEAEAVTVVSAI*

ATGGGTC AAGATGAGGTTGGGAGTGATCAGACGCAAATCATAAAGGGAAACGTACGAAG
CGACAAAGATCGTCTTCGACGTTTGTGGTGACGGCGGCACAACAGTGACTTCAACAAGT
TCATCGGCCGGTGGAAGTGGAGGAGAAAGAGCTGTTTCAGATGAATACAACCTCGGCCGTT
TCGTCTCCGGTGACTACTGATTGTACGCAAGAAGAAGAAGACATGGCGATTTGTCTCATC
ATGTTAGCTCGTGGGACAGTTCTTCATCGCCGGATCTCAAGAACTCGAGAAAAATT CAT
CAGAAGATTTTCGTCGGAGAATTCTAGTTTCTATGTGTACGAGTGTAAACGTGTAACCGG
ACGTTTTCGTGTTCCAAGCACTTGGTGGACACAGAGCGAGCCACAAGAAGCCGAGGACG
TCGACTGAGGAAAAGACACTAGACTACCTTGACGCAACCCAAAGTCTAGTGTCATCAGAAGT
GGGCAAAACAGTCATTTCAAAGTTTCCGGCTACGCCCTCAGCTTCAAGGCAAGTAAACAT
ATCAACAAGGCAACAAAGTACACGAGTGTTCCATCTGCGGTTCTGAGTTCACTTCCGGG

CAAGCTCTCGGTGGTCACATGAGGCGGCACAGGACAGCCGTAACCACGATTAGCCCCGTT
GCAGCCACCGCAGAAGTAAGCAGAAACAGTACAGAGGAAGAGATTGAGATCAATATAGGC
CGTTCGATGGAACAGCAGAGGAAATATCTACCGTTGGATCTTAATCTACCAGCACCAGAA
GATGATCTAAGAGAGTCAAAGTTTCAAGGGATAGTATTCTCAGCAACACCAGCGTTAATA
GATTGTCACTACTAG

>G358 Amino Acid Sequence (domain in AA coordinates: 124-135, 188-210)
MGQDEVGSDQTQIIKGRKTRQRSSSTFVVTAAATTVTSTSSAGSGGERAVSDEYNSAV
SSPVTDDCTQEEEDMAICLIMLARGTVLPSPDLKNSRKIHQKISSENSSFVVECKTCNR
TFSSFQALGGHRASHKKPRTSTEEKTRLPLTPKSSASEEQNSHFKVSGSALASQASNI
INKANKVHECSICGSEFTSGQALGGHMRRHRTAVTTISPVAATAEVSERNSTEEIEINIG
RSMEQQRKYLPLDLNLPAPEDDLRESKFQGI VFSATPALIDCHY*

>G360 (1..543)
ATGTGGAACCCCTAACAAATTGAAGAATTGGAGGATGATGATGAATCTTGGGAAGTCAAA
GCCTTTGAGCAAGACACTAAAGCAACATCTCTGGTACCACTTGGCCCTCCAAGATCTTAC
ACTTGCAATTTCTGCCGCCGTGAGTTCGGTCTGCTCAAGCCTTAGGCGGTACATGAAT
GTCCACCGCCGTGACCGCGCCTCATCTAGGGCTCATCAAGGTTCCACCGTTGCCGGCTGCC
GCTAGAAGCGGCCACGGGGGATGTTACTCAATTCTTGTGCTCCGCCGTTGCCCTACAACG
ACACTTATAATACAATCCACGCGAGTAACATTGAAGGTTTGTCCATTCTACCAACTG
CAAAACCCTAGTGGCATTTTTGGTAATTCTGGTGACATGGTGAATCTTTATGTAGAAGTT
CTCCTCGGCTTATTGAATATTCGACAGGAGATGATGAGAGCATTGGCTCGATGAAAGAA
GCGACAGGAACATCAGTGGATGAGCTTGATCTTGAACCTCGGCTAGGGCACCATCCACCG
TGA

>G360 Amino Acid Sequence (domain in aa coordinates: 42-62)
MWNPNKIEELEDDESWEVKAFEQDTKGNISGTTWPPRSYTCNFCRREFRSAQALGGHMN
VHRRDRASSRAHQGSTVAAAARSGHGGMLNSCAPPLPTTLIIQSTASNIEGLSHFYQL
QNPSGIFGNSGDMVNLYVEVPPRLIEYSTGDDESIGSMKEATGTSVDELDLELRLGHHP
*

>G362 (195..830)
ATAAAAAACCCCTTCATACAATATAAAATTTCTTTAGACATACAATATATTATACTATTAC
AGATGCAATGCATCATTAGTTACAAACTATTAAACTAAATATCCCCCGTCTCTCTCTTGC
TATATAAAGAAGATCATTACACATCTCCTTAAGCAAATTAAACCCATCGATAAACACAT
ACGTTACACATATATGTCTATAAATCCGACAATGTCTCGTACTGGCGAAAGTTCTTCAG
GTTTCGTCTCCGACAAGACGATAAAGCTATTTCGGCTTCGAACATCATCAGCGGCAGTCGTA
CGCCGGAAATCACGACGCGGAAAGCGTGAGCTCGTCCACAAACACGACGTCGTTAACAG
TGATGAAAAGACACGAGTGCCAATACTGCGGTAAAGAGTTTGCAAATTCTCAAGCCTTAG
GAGGTACCAAAAACGCTCACAAGAAGGAGAGGTTGAAGAAGAAGAGGCTTCAGCTTCAAG
CTCGCGAGCCAGCATCGGCTATTATCTACCAACCACCAACAACCCATAACGACGTCAT
TTCAGAGACAATACAAAACGCCGTCGTATTGTGCATTCTCTCCATGCACGTGAATAATG
ATCAGATGGGTGTGTACAACGAAGATTGGTTCGTGAGGTCGTGCGAGATTAACCTCGGTA
ATAATGACACGTGCCAAGATCTTAATGAACAAAGCGGTGAGATGGGTAAGCTGTACGGTG
TTCGACCGAACATGATTCAGTTCAGAGAGATCTGAGTTCGTCTGATCAGATGAGAA
GTATTAACCTCGCTGGATCTTCATCTAGGTTTTGCGGAGATGCGGCATAACAAATTAAG
AGAGATATATGATTAAGATTATATGTACTATAGTGGCGTATTTCAATTGGGATCATGAAGG
GGAAAAACGAGACATATAGTATTCTTGATGCAATTTGAGTTTTGTAATTTATTAGGTT
TATGTATGTTTTTCGAAG

>G362 Amino Acid Sequence (domain in AA coordinates: 62-82)
MSINPTMSRTGESSSSDKTIKLFGEFELISGSRTPETITAESVSSSTNTTSLTVMKRH
ECQYCGKEFANSQALGGHQNAHKERLKKRLQLQARRASIGYYLTNHQQPIITTSFQRQY
KTPSYCAFSSMHVNDQMGVYNEDWSSRSQINFNNDTQCQDLNEQSGEMGKLYGVRPNM
IQFQRLSSRSQMRINSLLDLHLGFAGDAA*

>G364 (64..516)
AAGCTTGATATCGCCTCTCTCTAATCTCTCTTCTCTCTCTATCTCTAAGAATATATAAA
GGTATGGACTACCAGCCAAACACATCCCTACGCTAAGCCTACCAAGTTACAAGAACCAC
CAACTAAACCTAGAAGTTGTTCTCGAGCCTTCTCCATGTCTTCTTCTTCTTCTTCC
ACGAACCTCATCATGTTTGGAGCAGCCTAGGGTATTCTCATGTAACCTATTGTCAAAGA
AAGTTTTACAGCTCTCAAGCTCTTGGTGGTCATCAAAACGCTCATAAGCTTGAGAGAACC
TTAGCCAAGAAGAGTCGAGAAGCTCTTTAGATCCTCAACACTGTTGATTCTGATCAGCCT

TACCCGTTCTCCGGTCGCTTTGAGCTTTACGGCCGTGGCTACCAAGGATTTCTCGAAAGT
GGCGGCTCGAGGACTTCTCCGCCCGCCGTGTGCCGAGAGTGGTCTTGATCAGGATCAG
GAGAAGAGTCACCTTGACTTATCCTTAAGGCTCTAAAAGAATCTTATATTTTGTAGTCT
ATATATTATCATATCAATTGTTAATCTTAAAATTGATTGTTTACTTATTAGTCATTTCC
TATTATCTGAAAGTTTTCTTTGTAAGTTGTAAGTATGGTCCTAAATCAAATCCAAATTT
GATTTTGGAAGATGGTACCTAATGCAGTAGTTAAATAAGTTAAAAAATGAAGGATCTAT
AATTCCT

>G364 Amino Acid Sequence (domain in AA coordinates: 54-76)
MDYQPNLSLRLSLPSYKNHQLNLELVLEPSSMSSSSSSSTNSSCLEQPRVFSCNYCQRK
FYSSQALGGHQNHLKRLTLAKSRELFSSNTVDSQPYPPFSGRFELYGRGYQGFLMSG
GSRDFSAARRVPESGLDQDQEKSHLDLSLRL*

>G365 (69..755)
CAATTCCTTTTACTTTCATTCTCTTTATATATTCTCTCTACGCTATAATATATATTACACA
GAATATACATGGACCGTCCATCAAAGGAGATCAAGAAATGTTAAAAATCAAGAAACAAG
GTCATCAAGATCTTGAGTTGGGGTTGACCCCTTTTGTCACGTGGAACCGCGACCTCATCAG
AGCTCAATCTCATCGATTCTTTCAAAACCAGCTCATCATCGACTTCTCATCATCAGCACC
AGCAAGAACAATTGGCAGATCCGAGAGTGTTCTCGTGTAATTATTGTCAAAGAAAGTTCT
ATAGTTCAACAAGCGCTAGGCGGTACCAAACGCTCATAAACGTGAGCGCACCTTAGCCA
AACGTGGACAGTATTACAAGATGACTCTCTCCTCCTTGCTTCTTCAGCGTTTGCGTTTG
GCCACGGTTCAGTCAGCAGATTCCGAAGCATGGCATCGTTACCATTACATGGCTCGGTGA
ATAACAGGTCAACGTTAGGGATTCAAGCTCATTCAACGATCCATAAGCCAGCTTCTTAG
GAAGACAAACGACGAGTTTAAGTCATGTTTTCAAACAGAGCATTACACCAGAAACCGACCA
TAGGAAAGATGTTGCCGAGAAATTCACCTTGAAGTCGCCGAAATAATAACAGTAACA
TGGTTGCTGCTAAGTTGGAGAGAATTGGACATTTCAAGAGCAACCAAGAAGATCATAATC
AGTTTAAGAAAATTGACTTGACTCTTAAGCTATGAGCTCTGCCATCTTCTTTTTAGTCTT
CATTATAACTTTTTTTATTCTCATCTTTGTTTGATATAATGATTGACGGCAGGGTGTT
AGAGTTTCACTAATGATCAAGTTGTACTTTTTATATATTTTCATTGATACCTTGTGATGT
AATCAATATTTTAGGTCTGTTTTT

>G365 Amino Acid Sequence (domain in aa coordinates: 70-90)
MEPSIKGDQEMLKIKKQGHQDLELGLTLLSRGTATSSSELNLIDSFKTSSSSSTSHHQHQQE
QLADPRVFSNVCYQKRFYSSQALGGHQNHLKRLTLAKRGQYKMTLSLSPSSAFAGHG
SVSRFASMASLPLHGSVNNRSTLGIQAHSTIHKPSFLGRQTTLSLHVFKQSIHQKPTIGK
MLPEKPHLEVAGNNNSNMVAALKERIGHFKSNQEDHNQFKKIDLTLLKL*

>G367 (1..708)
ATGGACGCTTCAATAGTTTCCTCATCCACTGCTTTTCCATATCAAGATTCTCTAAACCAG
AGCATCGAAGACGAAGAAAGAGACGTTTCATAATTCTAGTCACGAACCTCAATCTCATCGAC
TGCTAGACGACACAACGAGTATCGTTAACAATCTACAACATCCACAGAACAAGAGCTT
TTCTCATGCTACTATTGTCAAAGAACTTTCTATAGCTCACAAGCACTTGGTGGTCACCAA
AACGCACACAAGAGAGAGAACGTTGGCGAAGAGAGGACAACGTATGGCAGCGTCAGCC
TCAGCTTTTGACATCCTTACGGTTTCTCTCCACTTCCTTTCCACGGACAGTACAACAAC
CATAGGTCTTTAGGGATCCAAGCGCATTGATAAGCCACAAGCTAAGTTCTTATAACGGG
TTTGGTGGTCATATGGTCAGATCAACTGGTCAAGACTTCCATTGATCAACAACCAGCC
ATAGGTAAATTTCCCTCAATGGATAATTTTCATCATCATCATCATCAGATGATGATGATG
GCTCCTTCAGTAAATTCACGGTCCAATAACATCGATAGCCCAAGCAACACAGGACGGGT
CTAGAAGGGTCACCGACTCTTGAACAATGGCACGGAGACAAAGGATTGTTGTTAAGTACA
AGTCATCATGAAGAGCAGCAGAACTTGACTTGTCCCTCAAGCTTGA

>G367 Amino Acid Sequence (domain in AA coordinates: 63-84)
MDASIVSSSTAFPYQDSLNSIEDEERDVHNSSELNLIDCIDDTSIVNESTTSTEQKL
FSCNYQRTFYSSQALGGHQNHLKRLTLAKRGQRMASASAFGHPYGFSPLPFHGQYNN
HRSLGIAHSHSLKSSYNGFGGHYQINWSRLPFDQQAIGKFPSPMDNFHHHHHQMOMMM
APSVNSRSNNIDSPSNTGRVLEGSPTLEQWHGDKGLLLSTSHHEEQKLDSLKL*

>G373 (1..1854)
ATGGCGATTGAAACTCAGCTTCCTTGCGACGGTGACGGTGTGTATGCGGTGTCAGGTG
AATCCTCCGTGAGAAGAGACTCTCACTTGTGGCACGTGCGTCACTCCATGGCACGTGCCG
TGTCTCCTCCCCGAATCACTCGCTTCTTCCACTGGAGAGTGGGAGTGTCCCGATTGCTCC
GGCGTTGTCGTTCCCTCCGCCGCTCCGGGTACCGGAAACGCTCGACCTGAATCTCCGGT
TCAGTTCTCGTTGCTGCGATCCGTGCGATTCAAGCTGATGAGACTTTAACCGAAGCTGAG

AAAGCCAAAAAAGGCAGAACTGATGAGTGGGGGTGGTGACGATGGTGTGATGAAGAA
GAGAAGAAGAAGTTAGAAATCTTTTGTCTATTTCGATTCAATTGCCAGAAAGACCTATC
ACGACACCGTGTGGGCACAAATTTCTGTTTGAAATGTTTCGAGAAATGGGCAGTAGGTCAA
GGGAAGCTAACTTGTATGATATGCCGAAGCAAAATTCGAGACATGTGGCAAAAAATCCT
CGCATCAACTTAGCTCTAGTTTCTGCTATTTCGTTTAGCAATGTTACCAATGTTCTGTT
GAGGCAACTGCAGCCAAAGGTTTCATCATATTATCCGCAACCAAGACCGTCTGAGAAAGCA
TTTACTACCGAGCGGCAGTAAAACTGGGAAAGCTAATGCTGCTAGCGGTAAGTTTTTT
GTGACAATACCTCGTGATCATTTTGGTCCCATACCAGCTGAGAATGATGCTCACTAGAAAAG
CAAGGTGTTTTGGTTGGAGAATCTTGGGAGGACAGGCAAGAGTGTAGGCAGTGGGGAGCT
CATTTCCCGCATATTGCTGGCATTGCCGGCAATCAGCGGTTGGAGCTCAGTCTGTGGCC
CTCTCTGGAGGTTATGACGATGATGAGGATCATGGTGAATGGTTTTCTCTACACAGGAAGT
GGTGGGAAGGGATCTCAGTGGAAACAAAAGAATTAACAAGAAAACAGTCGCTGACCAGGCG
TTTAAAAACATGAATGAATCTCTAAGACTTAGTTGCAAAATGGGCTATCCTGTCCGAGTT
GTCAGTCTTTGGGAAGGAGAAGCGTTCTGCATATGCCCCCTGCTGAAGGTGTGAGATATGAT
GGGGTCTATCGAATTGAGAAGTGTGAGTAATGTTGGAGTACAGGGTCTTTTAAGGTC
TGTCGTTACCTGTTTGTAGATGTGACAATGAGCCAGCTCCATGGACCAGTGATGAGCAT
GGCGATCGTCCAAGACCGTTGCCTAATGTTCCGGAGCTTGAGACTGCTGCTGACCTGTTT
GTGAGAAAGGAGAGTCCATCATGGGATTTTCGATGAAGCTGAGGGTCTGTGGAAATGGATG
AAGTCTCCTCTGTTAGCAGAATGGCTTTGGATCCTGAGGAGAGGAAGAATAAGAGA
GCAAAAAATACATGAAGGCCAGACTTCTGAAAGAATTTAGTTGCCAAATCTGTCCGGAA
GTGCTGAGTCTTCCAGTGACGACGCCCTTGTGCACACAACCTCTGCAAGCATGCTTAGAA
GCGAAGTTTGTGGGATAACTCAACTGAGAGAGAGAAGCAATGGCGGACGTAAACTACGT
GCAAGAAGAACATCATGACCTGCCCTTGCTGCACGACGGATCTCTCCGAGTTTCTCCAA
AACC CGCAGGTGAACAGAGAGATGATGGAGATAATAGAGAATTTTAAGAAGAGTGAGGAA
GAGGCTGATGCATCCATTTCTGAAGAAGAAGAAGAATCCGAACCTCCAACCTAAGAAG
ATTAAGATGGATAACAACCTCTGTTGGTGGTAGTGGTACAAGTCTCTCAGCTTAA
>G373 Amino Acid Sequence (domain in AA coordinates: 129-168)
MAIETQLPCDGDGVCMRQVNPSEETLTCGTCVTPWHVPCLLPESLASSTGEWECPCDS
GVVVPSPAAPGTGNARPESGSLVLAIRAIQADETLTEAEKAKKRQKLMSSGGDDGVDEE
EKKKLEIFCSICIQLPERPITTPCGHNFCLKCFEKWAVGQGLTCMICRSKIPRHVAKNP
RINLALVSAIRLANVTKCSVEATAAKVHHIIRNQDRPEKAFTTTERAVKTGKANAASGKFF
VTIPRDHFGPIPAENDVTRKQGVLVGESWEDRQECROWGAHFPFIAGIAGQSAVGAQSV
LSGGYDDDEDHGEWFLYTGSGGRDLGNKRINKKQSSDQAFKNMNESSLRLSCKMGPVVRV
VRSWKEKRSAYAPAEGVRYDGVYRIEKCWSNVGVQGSFKVCRYLFVRCDNEPAPWTSDEH
GDRPRPLPNVPELETAADLFVRKESPSWDFDEAEGRWKWMKSPVSRMALDPEERKKNKR
AKNTMKARLLKEFSCQICREVLSPVTTCAHNFKACLEAKFAGITQLRERSNGGRKLR
AKKNIMTCCPCTTDLSEFLQNPQVNREMEIIEFKSEEBADASISEEEEESEPPTKK
IKMDNNSVGGSGTSLSA*
>G396 (1..957)
ATGGGGGAAAGAGATGATGGGTTGGGTTGAGTCTAAGCTTGGGAAATAGTCAACAAAAA
GAACCATCTCTGAGGTTGAATCTTATGCCGTTGACAACCTCTTCTTCTTCTTCTGTTT
CAACACATGCACAATCAGAATAACAATAGCCATCCCAGAAAGATTACATAACATCTCTTGG
ACTCATCTGTTTCAATCTTCTGGGATTAAACGTACAACCTGCAGAGAGAAACTCCGACGCC
GGGTCAATTCTAAGAGGTTTCAACGTGAACAGAGCTCAGTCTTCCGTGGCGGTAGTGGAC
TTGGAAGAAGAAGCCGCCGTCGTCTCGTCTCCAAACAGCGCCGTTTCGAGTCTGAGTGG
AATAAAAGGGATCTTGGCGTGGCGAGAGGAGAGATGAAAACGAGGCGGAGAGAGCTTCT
TGCTCACGCGGAGGGGGAAGCGGTAGCGACGATGAAGACGGCGGAACGGCGACGGA
TCAAGGAAGAACTACCGTTATCGAAGGATCAAGCTCTTGTCTCGAGGAGACTTTTAA
GAACATAGCACTCTTAATCCGAAGCAAAAGCTGGCTCTAGCAAAACAGTTGAATCTAAGG
GCAAGACAAGTTGAAGTGTGGTTTCAGAACCGTAGGGCAAGGACGAAGCTGAAACAAACG
GAGGTTGATTGTGAGTATTTAAAGAGATGTTGCGATAATCTGACCGAGGAGAATCGACGG
CTGCAGAAAGAAGTGTGCGAGCTGAGGGCGTTGAAGTTGTCTCCACATCTCTACATGCAC
ATGACTCCTCTACTACTCTCACCATGTGCCCTTCTTGCGAACGTGTCTCCTCCTCTGCC
GCCACTGTGACCGCTGCTCCTTCCACTACTACTACTCCTACGGTGGTGGGGCGGCCAAGT
CCACAGCGATTAACTCCTTGGACTGCTATTTCTCTCCAGCAAAAATCAGGTGCTAG
>G396 Amino Acid Sequence (domain in AA coordinates: 159-220)
MGERDDGLGLSLSLGNSQQKEPSLRNLMLPLTSSSSSFQHMHNQNNNSHPQKIHNISW

ATGGAGAGTGGTTCCAACAGCACTTCTTGTCCAATGGCTTTTGCCGGGGATAATAGTGTAGGTCCGATGTGTCCATGATGATGATGATGCCGCCCATCATGACATCACATCAACATCATGGTCATGATCATCAACATCAACAACAAGAACATGATGGTTATGCATATCAGTCACACCACAACAAAGTAGTTCCTCTTTTCTTCAATCACTAGTCTCCTCCCAAGGAACTAAGAACAAGTTGCTTCTTCTTCTTCTCTCTCTCTTGTGCTCCTGCCTATTCTCTAATGGAGATCCATCATAACGAAATCGTTGCAGGAGGAATCAACCCCTTGCTCCTCTTCTCTTCTTCAGCCTCTGTCAAGGCCAAGATCATGGCTCATCCTCACTACCACCGCCTCTTGCGCGCTTATGTCAATTGTCAAGAGTTGGAGACACCACCGAGGTTGTGGCGAGGCTGGAGGAGGCATGCTCGTCTGCCGAGCGCAGCGCATCTATGGGGCCAACAGGGTGCTTGGTGAAGATCCAGGGCTTGATCAATTGATGGAAGCTTACTGTGAAATGCTCGTTAAGTATGAGCAAGAGCTCTCCAAACCTTTCAAGGAAGCTATGGTCTTCCCTTCAACGTGTCGAGTGTCAAATCAAATCCCTCTCTCTATCCTCACCTTCTCTTTCTCCGGTTATGGAGAGACAGCAATTGATAGGAACAATAATGGGTATCCGAGGAAGAAGTCGATATGAACAATGAATTTGTAGATCCACAAGCTGAGGATAGAGAGCTTAAAGGACAGCTCTTGCGCAAGTACAGTGGTTACTTtagggagcctcaagcaaGAGTTTCATGAAGAAGAGGAAGAAAGGAAAGCTCCCTAAAGAAGCTCGTCAACAACTGCTTGATTGGTGGAGCCGTCCTACTACAAATGGCCTTACCCTTCGGAGCAACAAAAGCTCGCCCTTGCGGAATCAACGGGGCTGGACCAGAAAAGATAAACAATTGGTTTCATAAACCAGAGGAAACGGCATTTGGAAGCCGTCGGAGGACATGCAGTTTGTAGTAATGGACGCAACACATCCTCACCTACTTTCATGGATAATGTCTTGGACAATCCTTCCCAATGGATCACATCTCTCCACCATTGCTTTGA

>G431 Amino Acid Sequence (domain in AA coordinates: 286-335)
MESGSNSTSCPMAFAGDNSDGPMPMMMMPPIMTSHQHGHGHDHQHQQQEHDGYAYQSHH
QQSSSLFLQSLAPPQGTKNKVASSSSPSSCAPAYSLMEIHNEIVAGGINPCSSFSSAS
VKAKIMAHPHYHRLLAAYVNCQKVGAPPEVVARLEEACSSAAAAAASMGPTGCLGEDPGL
DQFMEAYCEMLVKYEQELSKPFKEAMVFLQRVECQFKSLSLSSPSSFSGYGETAIDRNNN
GSSEEVDMNNEFVDPQAEDELKQGLLRKYSGYLGLSKQEFMKKRKKKGKLPKEARQQLL
DWFSRDYMKWYPSPQQKLALAEKSTGLDQKQINNWFINQYKRKHWPSPEDMQFVVM DATHPH
HYFRMNVLDNPPEMDHSLASTL*

ATGGAGATGGGGTTCCTCAACTCGGGTCCGGGTCATGGTCCGGGTCAAGCAGAGTCGGGTGGT
TCCTCCACTGAGTCATCCTCTTTCACTGGAGGGCTCATGTTTGCCAGAAAGATCTACTTC
GAGGACGGTGGTGGTGGATCCGGGCTCTCTTCCTCAGGTGGTCGTTCAAACAGACGTGTCT
CGTGGAGGCGGGTCCGGTTCAGTCGGGTTCAGATACCAAGGTGCCAAGTGGAAAGGTTGTGGG
ATGGATCTAAACCAATGCAAAAGGTTATTACTCGAGACACCGAGTTTGTGGAGTGCACCTCT
AAAAACCTAAAGTCACTGTGGCTGGTATCGAACAGAGGTTTTGTCAACAGTGCAGCAGG
TTTCACTCAGCTTCCGGAATTTGACCTAGAGAAAAGGAGTGGCCGAGGAGACTCGTGGT
CATATGAGCGCAGCAAGGAAGGCCACCGTCGCTCTCTCTCTGTGTAGCTTCTCGTTAC
GGGAGGATCGCACCTTTCGCTTACGAAAATGGTGATGCTGGAAAGAATGGAAGCTTTCTT
GGGAACCAAGAGATAGGATGGCCAAGTTCAAGAACATTGGATACAAGAGTGATGAGGCGG
CCAGTGTCTCACCGTCATGGCAGATCAATCCAATGAATGTATTTAGTCAAGGTTCAAGTT
GGTGGAGGAGGGACAAGCTTCTCATCTCCAGAGATTATGGACACTAACTAGAGAGCTAC
AAGGGAATTGGCGACTCAAACCTGTGCTCTCTCTCTCTGTCAAATCCACATCAACCACAT
GACAACAACAACAACAACAACAACAGCAACAACAACAACAATACATGGCGAGCTTCT
TCAGGTTTTGGCCCGATGACGGTTACAATGGCTCAACCACCACCTGCACCTAGCCAGCAT
CAGTATCTGAACCCGCCCTTGGGTATTCAAGGACAATGATAATGATATGCTCTCTGTTTTG
AATTAGCTCGATACACCGAGCCAGATAATTGTGAGATAAGTAGTGCCACGGCAATGGGT
GAGTTTCAGGTTATCTGATCACCATCATCAAAGTAGGAGACAGTACATGGAAGATGAGAAC
ACAAGGGCTTATGACTCTTCTTCTCACCATAACCACTGGTCTCTCTGA

MEMGSNSGPGHGPGQAESGGSSSTESSFSGGLMFGQKIYFEDGGGGSGSSSSGGRSNRRV
RGGGSGSGGOIPRCQVEGCGMDLTNAKGYYSRHRVCGVHSKTPKVTVAGIEORFCOOCSE

FHQLPEFDLEKRSRRRLAGHNERRRKPQPASLSVLASRYGRIAPSLYENG DAGMNGSFL
GNQEIGWPSSRTLDTRVMRRPVSSPSWQINPMNVFSQGSVGGGTSFSSPEIMDTKLESY
KGIGDSNCALSLLSNPHQPHDNNNNNNNNNNNTWRASSGFGPMTVTMAQPPAPPSQH
QYLNPPWVFKDNDNDMS PVLNLGRYTEPDNCQISSGTAMGEFELSDHHHQSRRQYMEDEN
TRAYDSSSHHTNWSL*

>G546 (1..588)

atgactcgaccgtcaagattacttgagacggcgccgaccaccacacacgcgcgaggag
atgatcgacgcgaatccgacatgggtggatcttgctcggtctcttcttgctctctatc
tgcgttgctggtctcgccgcgctcgtagcatgcgcttggtccggcggtttacagccgga
ggagattcgccgctcaccgaacaaaggcttgaaaaagaaagctcttcagttcttccaaga
tccactttcaccgcgcggaatcaacctccggcgccgcgctgaagaggagactcgacg
gaatgtgctatttgctcactgacttcgcccgcggtgaagaaataagagtgtcttctctt
tgtgtgctattcttccacgtggagtgtattgacaaatggctagtttctaggtcttctgt
cttcttctgctcaggattcttaccgcggtgagatgtgaccggtgtggtcatgtctctacg
gcggagatgaagatcaagctcatcgatcatcaacatcaccaacactcttctactaccatt
cctacgtttcttcttcttaa

>G546 Amino Acid Sequence (domain in AA coordinates:114-155)

MTRPSRLLETAAPPPQPPSEEMIAAESDMVILSALLCALICVAGLAADVRCALRRFTAG
GDSPPSNKGLKKALQSLPRSTFTAAESTSGAAEEGDSTECALCLTDFADGEEIRVLPL
CGHSFHVCEIDKWLVSRSRSCPSRRILTPVRCDRCGHASTAEMKDQHRHQHHQSSTTI
PTFLP*

>G551 (1..708)

ATGGAGTGGTCAACAACGAGCAACGTAGAAAACGTGAGAGTAGCTTTTCATGCCACCGCCA
TGGCCGGAGTCTAGTTCCTTTAACTCGCTCCACAGCTTCAACTTGTATCCTTACGCAGGA
AATTCATATACGCCGTGGCGATACACAAACCGGACCGGTTATCTCTGTACCGGAATCAGAA
AAGATCATGAATGCGTACCGATTTCGGAACAACAACATGAGATGATAAAAAAGAAGAGA
CTAACGAGTGGACAATTAGCTTTCATTTGAGCGAAGTTTCAAGAAGAGATCAAATTAGAT
TCAGACAGGAAGGTGAAGCTGTGAGAGAGCTCGGTCTGCAGCCACGTCAGATAGCAGTT
TGGTTCCAAAACCGCGTGCACGGTGAAGGCGAAGCAGCTTGAGCAGTTGTACGACTCG
CTTAGACAAGAGTACGACGTCGTTTCTAGGGAGAAACAAATGTTACACGATGAGGTGAAG
AAGCTGAGAGCTTTACTAAGAGACCAGGGTTTGATCAAGAAGCAAAATCTCTGCCGGGACC
ATCAAAGTTTCCGGTGAGGAAGACACGGTGGAGATTTTCATCGGTGGTGGTAGCTCATCCA
AGAACGGAGAATATGAACGCAAAATCAAATCACCGGAGGGAATCAAGTTTACGGTCAATAC
AACAATCCGATGCTGGTTGCTTCTCTGGCTGGCCGTCATACCCCTGA

>G551 Amino Acid Sequence (conserved domain in AA coordinates:73-133)

MEWSTTSNVENVRAFMPPPPWPESSSFNSLHSFNFDPYAGNSYTPGDTQTGPVISVPESE
KIMNAYRFPNNNNEMIKKKRLTSGQLASLERSFQEEIKLSDSRKVKLSRELGLQPRQIAV
WFQNRARRWAKQLEQLYDSLROEYDVVSREKQMLHDEVKKLRALLRDQGLIKKQISAGT
IKVSGEEDTVEISSVVVAHPRTENMNANQITGGNQVYGQYNNPMLVASSGWPSYP*

>G578 (1..978)

ATGCATAGTTTGAATGAAACAGTAATTCCTGATGTTGATTACATGCAGTCTGATAGAGGG
CATATGCATGCTGCTGCTCTGATTCCAGTGATCGATCAAAGGATAAGTTGGATCAAAAG
ACCCCTTCGTAGGCTTGCTCAAATCGTGAGGCAGCAAGAAAAAGCAGATTGAGGAAGAAG
GCGTATGTTGAGCAGCTGGAAGATAGTCGATTAAAGCTGACTCAAGTTGAGCAGGAGCTG
CAAAGAGCAAGACAGCAGGGAGTTTTCATCTCAAGTTGAGGAGACCAAGCTCATTCTACT
GGTGGCAATGGTGGGGCTTTGGCATTGATGCAGAACACTCACGATGGCTTGAAGAAAAG
AACAGGCAAAATGAACGAGCTGAGATCTGCCCTGAATGCTCATGCAGGTGATACTGAGCTC
CGGATAATTGTGGATGGAGTGATGGCTCACTATGAGGAGCTTTTCAGGATTAAGAGCAAT
GCATCTAAGAAATGATGTCTTCCACTTGTATCTGGAATGTGGAAAACACCAGCTGAGCGA
TGTTTCTTGTGGCTTGGCGGGTCCCGTCATCCGAACCTTCTCAAGCTTCTTGGCAATCAG
CTAGAGCCCATGACAGAACGACAGGTAATGGGCATCAATAGCTTGCAGCAGAGCTCGCAG
CAGGCAGAGATGCTTTTATCTCAAGGGATGGAGAGTTTACAGCAATCCCTAGCTGATACT
TTATCCAGTGAACCTCTTGGTTCCAGTTTCATCGGATAATGTGCGGAGCTACATGGGTCAG
ATGGCCATGGCAATGGGCAAGTTAGGCACCCTCGAAGGATTCATACGCCAGGCTGATAAC
TTGAGGCTGCAAACTACAAACAGATGCTTCGAGTATTAACAACACGTCAGTCAGCTCGT
GCTCTTCTTGTATACACGATTATTCATCTCGATTACGTGCTCTTAGTTCCTTGTGGCTT
GCCCGGCAAGAGAGTGA

>G578 Amino Acid Sequence (domain in AA coordinates 36-96)
MHSLNETVIPDVDMQSDRGHMHAAASDSSDRSKDKLDQKTLRRLAQNREAARKSRLRKK
AYVQQLSRLKLTQVEQELQRRARQQGVFISSSGDQAHSTGGNGGALAFDAHSRWLEEK
NRQMNELRSALNAHAGDTLRIIVDGVMAHYEELFRIKSNASKNDVFHLLSGMWKTPAER
CFLWLGGFPSELLKLLANQLEPMTERQVMGINSLQQTSSQQAEDALSQGMESLQQSLADT
LSSGTLGSSSSDNVASYMGQMAMAMGKLTLEGFIRQADNLRQLQLQMLRVLTTTRQSAR
ALLAIHDYSSRLRALSSLWLARPRE*

>G596 (168..1121)

TAATTTCTCTACTTCAGATTTTTTCTCCTTAGATTAATTTAATTGAGTTATTGTACATC
CCTCAAGCTAAGATTCTGTTTTGTGAGTTGAGTGGATGAGAAGAGGAGAGATTAATAA
ATTAGGGTTCAATTGTTTACTTTTTGTGTTTTTATATCAAGTAATGGATCAGGTCT
CTCGCTCTCTCTCCACCTTTTCTCTCAAGAGATCTCCATCTTCACCCACACCATCAAT
TCCAGCATCAGCAGCAGCAGCAACAGAATCACGGCCACGATATAGACCAGCACCGAA
TCGGTGGGCTAAAACGTGACCGAGATGCTGATATCGATCCCAACGAGCACTCTTCAGCCG
GAAAAGATCAAAGTACTCCTGGCTCCGGTGGAGAAAGCGGCGGAGGAGGAGGAGATA
ATCACATCACGAGAAGGCCACGTGGCAGACCAGCGGGATCTAAGAACAACCAAAACCGC
CAATCATCATCTACTCGAGACAGCGCAACGCTCTCAATCTCATGTTCATGGAAGTAGCAA
ACGGATGTGACGTCTGGAAGTGTACCGTCTTCGCTCGCCGTCGCCAACGTGGCATCT
GCGTTTTGAGCGGAAACGGCGCGTTACCAACGTTACCATAAGACAACAGCTTCAGTAC
CTGGTGGTGGCTCATCTGTCTTAACCTACACGACGTTTCGAGATTCTTTCTCTCTCGG
GATCATTCCTCTCTCCGCTCCACAGCTGCGTCAGGTCTAACGATTTACTTAGCCG
GTGGTCAGGGACAGGTTGTTGGAGGAAGCGTGGTGGTCCACTCATGGCTTCAGGACCTG
TAGTGATTATGGCAGCTTCGTTTGGAAACGCTGCGTATGAGAGACTGCCGTTGGAGGAAG
ACGATCAAGAAGAGCAAAACAGCTGGAGCGGTTGCTAATAATATCGATGGAAACGCAACAA
TGGGTGGTGGAAACGCAACGCAAACTCAGACGACAGCAACAGCAACAGTTGATGC
AAGATCCGACCTCGTTTATACAAGGGTTCCTCCGAATCTTATGAATTCTGTTCAATTGC
CAGCTGAAGCTTATTGGGGAACCTCCGAGACCATCTTTCTAAATCGCGAAGAAAAACAAG
TTAGATACGTTTCGTTGTTTTTAATTTATAATCTCTCTTCTGTCAAGTTTTAATTTCTTT
TTCTTCTTCTTTGTTTTCTAAAGATAATTGTAAGTCTTTGACGAAGATTCGTGGTACGTAT
GAATCGAAGAGAATCGTTTTGTCATGGGATTGCTCGATCTATTAGGTTTGAGAGGGGGT
TTGTGTTTTGCGTTGACTAGCAGATTATAAAATTGTTGATTTTCGAGTTTTTATTTTCAT
GTGTTGGTGATAAA

>G596 Amino Acid Sequence (domain in AA coordinates: 89-96)

MDQVSRSLPPPLSRDLHLHPHQFQHQQQQQNHGHDIDQHRIGGLKRDRDADIDPNE
HSSAGKDQSTPGSGGESGGGGGDNHITRRPRGRPAGSKNPKPPIIITRDSANALKSHV
MEVANGCDVMESVTVFARRRQRGICVLSGNGAVTNVTIRQPASVPVGGSSVVNLHGRFEI
LSLSGSFLLPPAPPAAASGLTIYLAGGQGVVGGSVVGPLMASGPVIMAASFGNAAYERL
PLEEDDQEEQTAGAVANNIDGNATMGGGTQTQTQTQQQQQQLMQDPTSFQGLPPNLMN
SVQLPAEAYWGTTPRPSF*

>G617 (59..1141)

CAGATCTGTTCTTTACACCAAATTGAGTACTGAAGATCTTGTGAGTGAATTAAAGAGAT
GAGATCAGGAGAATGTGATGAAGAGGAGATTCAAGCAAAGCAAGAAAGAGATCAAAATCA
AAATCATCAAGTAAACTTAAACCACATGTTGCAACAACAACAGCCGAGTTCCGTATCATC
TTCAAGGCAATGGACTTCAGCTTTTAGGAATCCAAGAATCGTTCGAGTCTCAAGAACATT
CGGTGGCAAAGACAGACACAGCAAAGTATGTACAGTCCGTGTTTCGAGACCGGAGGAT
AAGGTTGTCCGTACCTACAGCTATTCAACTCTACGACCTTCAAGATCGATTAGGGCTGAG
TCAGCCAAGCAAAGTCATTGATTGGTTACTCGAAGCAGCAAAAGATGACGTAGACAAGCT
ACCTCCTCTACAATCCACATGGATTTAACCAGATGTATCAAATCTCATCTTCGGAAA
CTCCGGGTTTGGAGAATCTCCATCTTCAACTACATCAACAACGTTTCCAGGAACCAATCT
CGGGTTCTTGGAAAATTGGGATCTTGGTGGTTCTTCAAGAACAAGAGCAAGATTAACCGA
TACAACCTACGACCAAGAGAAAGTTTGGATCTTGATAAAGGAAAATGGATCAAAAACGA
CGAGAATAGTAATCAAGATCATCAAGGGTTTAAACACCAATCATCAACAACAAATTCCTCT
GACCAATCCGTACAACAACACTTCAGCTTATTACAACCTTGGACATCTTCAACAATCGTT
AGACCAATCTGGTAATAACGTTACTGTGCAATATCTAATGTTGCTGCTAATAATAACAA
TAATCTCAATTTGCATCCTCTCTCCTCGTCTGCCGAGATGGATCTCAGCTTTTTTTCGG
TCCTACTCCTCCGCAATGAGCTCTCTATTCCCGACATACCTTCGTTTCTTGGAGCTTC
TCATCATCATCATGTCGTGATGGAGCCGGTCATCTTCAGCTCTTTAGCTCGAATTCAAA

TACCGCATCGCAGCAACACATGATGCCGGGTAATACGAGTTTGATTAGACCATTTCATCA
TTTGATGAGCTCGAATCATGATACGGATCATCATAGTAGCGATAATGAATCAGATTCTTG
AATGATTTTATATATCTACACTATACATTGAAAAATGTTATATGTATACGTATTCTTCTAT
ATTTTGATATATATGCGTATTGTTGGATTGGTTTATGTATCT

>G617 Amino Acid Sequence (domain in AA coordinates: 64-118)
MRSGECDEEEIQAKQERDQNNHQNHLNMLQQQPPSSVSSSRQWTSAFRNPRIVRVSRT
FGGKDRHSKVCTVRGLRDRIRLSVPTAIQLYDLQDRLGLSQPSKVIDWLLEAAKDDVDK
LPPLQFPHGFNQMYPNLIFGNSGFGESPSSTTSTTFPGTNLGFLENWDLGGSSRTRARLT
DTTTTQRESFDLDKGKWKINDENSNDHQGFNTNHQQQFPLTNPYNNTSAYYNLGHLLQS
LDQSGNNVTVAISNVAANNNNNLNLHPPSSSAGDGSQLEFFGPTPPAMSSLFPTYPSTFLGA
SHHHVVDGAGHLQLFSSNSNTASQQHMMPGNTSLIRPFHMLMSSNHDTHHSSDNESDS
*

>G620 (40..666)
GAATTGAACCTGGACCAGCACAGCAACAACCCCAACCCCAATGACCAGCTCAGTCATAGTA
GCCGGCGCCGGTGACAAGAACAATGGTATCGTGGTCCAGCAGCAACCACCATGTGTGGCT
CGTGAGCAAGACCAATACATGCCAATCGCAAACGTGATAAGAATCATGCGTAAACCTTA
CCGTCTCACGCCAAAATCTCTGACGACGCCAAAGAAACGATTCAAGAATGTGTCTCCGAG
TACATCAGCTTCGTGACCGGTGAAGCCAACGAGCGTTGCCAACGTGAGCAACGTAAGACC
ATAACTGCTGAAGATATCCTTTGGGCTATGAGCAAGCTTGGGTTGATAACTACGTGGAC
CCCCTCACCGTGTTCATTAAACCGGTACCGTGAGATAGAGACCGATCGTGGTCTGCACTT
AGAGGTGAGCCACCGTCGTTGAGACAAACCTATGGAGGAAATGGTATTGGGTTTCACGGC
CCATCTCATGGCCTACCTCCTCCGGTCTTATGGTTATGGTATGTTGGACCAATCCATG
GTTATGGGAGGTGGTCCGTACTACCAAACCGGTCGTCGGGTCAAGATGAATCCAGTGT
GGTGGTGGCTCTTCGTCTTCCATTAAACGGAATGCCGGCTTTTGACCATTATGGTCAGTAT
AAGTGAAGAAGGAGTTATCTTCTTATTTTATATCTATTCAAAACATGTGTTTCGATAGAT
ATTTTATTTTATGTCTTATCAATAACATTTCTATATAATGTTGCTTCTTTAAGGAAAAG
TGTTGTATGTCAATACTTTATGAGAACTGATTATATATGCAAT

>G620 Amino Acid Sequence (domain in AA coordinates: 20-118)
MTSSVIVAGAGDKNGIIVVQQPPCVAREQDQYMPIANVIRIMRKTLP SHAKISDDAKET
IQECVSEYISFVTGEANERCQREQRKTITAEDILWAMSKLGFNDYVDPLTVFINRYREIE
TDRGSALRGEPPSLRQTYGGNGIGFHPGSHGLPPPGPYGYGMLDQSMVMGGGRYYQNGSS
GQDESSVGGSSSSSINGMPAFDHYGQYK*

>G625 (151..1137)
AATCGACCATTCACAACGATGACATTCAAACACTCTTCAGTTTCCCTTCTCTTCTTGATTC
GTCCTCTCCACTATTTTCTCAATTTCTTTAATCTCTCTTTCTCTCTCTACTTCTCT
TCCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
AACCATCTGGAAGATAATAACCAAACCTAACCATAATAATCTCAATCCGATTCCACC
ACCGACTCATCACTTCTTCCGCTCAACGCAAACGCAAAGGCAAAGGTGGTCCGGACAAC
TCCAAGTTCCGTTACCGTGGCGTTTCGACAAAGAAGCTGGGGCAAATGGGTCCGCCGAGATC
CGAGAGCCACGTAAGCGCACTCGCAAGTGGCTTGGTACTTTCGCAACCGCCGAAGACGCC
GCACGTGCCCTACGACCGGGCTGCCGTTTACCTATACGGGTACGTGCTCAGCTCAACTTA
ACCCCTTCTGCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
CCTTCCACCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
GCCACCGTAGGAGGAGGACCAACTTTGGTCCGTACGGTATCCCTTTTAAACAACAACATC
TTCCTTAATGGTGGGACCTCTATGTTATGCCCTAGTTATGGTTTTTTCCTCAACAACAA
CAACAACAAAATCAGATGGTCCAGATGGGACAATTCCAACACCAACAGTATCAGAATCTT
CATTCTAATACTAACAATAACAAGATTTCTGACATCGAGCTCACTGATGTTCCGGTAACT
AATTCGACTTCGTTTCATCATGAGGTGGCGTTAGGGCAGGAACAAGGAGGAAGTGGGTGT
AATAATAATAGTTCGATGGAGGATTGAACCTCTTAGCTGGTTCGGTGGGTTCGAGTCTA
TCAATAACTCATCCACCGCGTGGTGGTATCCGGTATGTTCTATGGGTCTGGATCCGGGT
TATATGGTGGAGATGGATCTTCGACCATTTGGCCTTTTGGAGGAGAAGAATAATAGT
CATAATTGGGGGAGTATTTGGGATTTTATTGATCCCATCTGGGGGAATTTCTATTAAATTT
GTTTTTGTGGAAGATCATATTATATACGATGAGCATCCCTAAGGTCGGTCAAGAGCATTG
GAGATTCAATTGTTGAGAGGAATCAAAGAGATTGCATTCTATGAGGAGCTCTGCATGCAAA
ATTTTGGAGGATTTTCTTACTACCTATAGAGATAAATAAGAGGGTATTTTATTATTTT
TTGAAGATTTTATTTTCAAGGAATTCGTAAAAGAGATTACGGTTCCAATAAAGTATGTA
TATGTGAAGAGAATCGGAGGAGATGGTGGAAAGTTGTATGGGAATTTTATTGGTTCAAC

ACTCCTCCTCCTCCAACGAATTGTAGCTATAGGTTGTTGGATTGATCTCACAAGCAAT
TCTCCTGCTCAATCCCTCAAGACAAGCAACCGATGGATACTTGTGGAGCTGCCAAGTGT
CAAGAACCCATCACTCCAACCTCAATGAGTGAGCAGAAGAAGCAACAAACATCAAGAAGT
CGAATAAAGTGCAAATGCAAGGCATTGCGGTTGGTTCGTGCGGTTGATTTAACTGTTG
AAATCTTACGATGAAGTGAATGATGAGCTTGAGGAGATGTTTGAAGATTCAAGGACAGCTT
CTTGCCCGAGACAAATGGATCGTTGTCTTCACTGATGATGAAGGAGATATGATGCTTGCT
GGTGATGATCCGTGGAATGAGTTTTCAGATGGCAAAGAAGATATTTATATATTTCGAGC
GATGAGGTTAAGAAATGACAACGAACTGAAGATTTCTTCGTGTTAGAGAATGAGGAA
TATGGTAATGAATCATTGCAAAATCGTAGTAGGGGTGAGAGTTTTAGCTGTTAATTAAG
GTTAATTCGCGCAGCTCGTTTTAGTGCGTAAGTGCTAAAGACTTTTTTTTAGTCTGTG
TATATAAAGTCTTGCTCTTTTTTCATGTCAATTTTCAAGTTGGCGATTAAATATTTTCG
GTTTTGGGACAGTGGTTGATGGGGCGGTTTTACATTTTTTATGTGTATGTAATTTGCCA
AAACCATTCATTTTCAAA

>G716 Amino Acid Sequence (domain in AA coordinates: 24-355)
MASVEGDDDFGSSSSRSYQDQLYTELWKVCAGPLVEVPRAQERVVFYFPGHMEQLVASTN
QGINSSEIPVFDLPKILCRVLDVTLKAEHETDEVYAQITLQPEEDQSEPTSLDPPIVGP
TKQEFHSFVKILTASDSTHGGFSVLRKHATECLPSLDMTQATPTQELVTRDLHGFEWRF
KHIFRQPRRHLLTTGWSTFVSSKRLVAGDAFVFLRGENDLRVGVRLARHQSTMPTSV
ISSQSMHLGLVLTASHAVRTTTFVVFYKPRISQFIVGVNKYMEAIKHGFSLGTRFRMRF
EGEESPERIFTGTIVGSGDLSSQWPASKWRSLOVQWDEPTTVQRPDKVSPWEIEPFLATS
PISTPAQQPQSKCKRSRPIEPSVKTPAPPSFLYSLPQSQDSINASLKLFDPSLERISGG
YSSNNSFKPETPPPTNCSYRLFGLDLSNSPAPIPDQKQPMDCGAACQEPITPTSMS
EQKKQOTSRSRTKVQMGIAGRAVDLTLKSYDELIDELEEMFEIQGQLLARDKWIVVF
TDDEGDMMLAGDDPWNEFCMAKKIFIYSSDEVKMTTKLKISSLENEEYGNESFENRS
RG*

>G725 (46..1122)
CCTCTTTTCAGAGAGAGAAAGAGAGTCAGAGAGAGAGAGAGAGAGAAATGTTCCATGCTAAG
AAACCTTCAAGTATGAATGGTTCATATGAGAACAGAGCTATGTGCGTTCAAGGCGATTCA
GGCCTTGTCTCACCACCGACCTAAACCGGTTTGGCGTTGGACCGTCGAACTCCACGAG
CGTTTTGTGACGCGCTCGCTCAGCTCGGCGGCCCCGACAAAGCGACCCCAAGACGATT
ATGAGAGTTATGGGTGTGAAGGTTCTACTCTTTACCACCTAAAGAGCCATCTTCAGAAA
TTCAGGCTTGGAAAGCAGCCGACAGAGAGTACGAGATCACTCCACAAAGGAAGGTTCA
AGAGCTTCTGCCATGGATATTTCAGCGCAACGTAGCTTCTTCTTCTGCGATGATGAGTCGC
AACATGAATGAGATGCAAAATGGAAGTGCAGAGAAGGTTGCATGAACAGCTAGAGGTGCAA
AGACATCTGCAACTGAGGATTGAAGCACAAGGAAAGTACATGCAATCTATCTTGGAGAGA
GCTTGCCAAACCTAGCCGGTGAGAACATGGCAGCCGCCACCGCAGCAGCCCGCTCGGA
GGAGGATACAAGGTAATCTGGGAAGTTCGAGTCTTTCAGCAGCGGTGGGCCCCACCTCCT
CATCCTCTTAGTTTTCCCGCGTTTCAAGACCTAAACATCTATGGAACACAACCGACCAA
GTCTTCGACCATCAAACTTCCATCATCAAAACATAGAGAACCATTTACGGGTAACAAT
GCTGCAGACACCAACATTTACTTGGGGAAGAAGCGACCTAATCCTAATTTTGGTAACGAT
GTAAGGAAAGGACTATTGATGTGGTCTGATCAAGATCAGATCTTCCGCAAACCAATCG
ATCGATGATGAGCATAGAATTCAGATACAGATGGCTACACATGTCTCCACGATTGGAT
TCTTTGTGCGAGATCTACGAAAGGAAATCAGGTTTATCAGGTGATGAAGGGAATAATGGT
GGGAAATTACTGGAAAGGCCATCGCCTAGGAGATCACCATTGAGTCTTATGATGAACCTT
AATGGTGGATTAATACAAGGAAGAACTCGCCATTTGGGTGATACAATTTATTAATTTT
ATCTATGAGTGATGCATGGGAATGTAAGAACGAGATATATATGTTTGTCTATTGTGAGTT
TGACGTAGGTTTATAGAGAAA

>G725 Amino Acid Sequence (domain in AA coordinates: 39-87)
MFHAKKPSSMNGSYENRAMCVQDGLVLTDPKPRLRWTVELHERFVDAVAQLGGPDKA
TPKTIMRVMGVKGLTLYHLKSHLQKFLGKQPHKEYGDHSTKEGSRASAMDIQRNVASSS
GMMSRNMNMQMEVQRRLEHQLVQRHLQLRIEAQKYMQSILERACQTLAGENMAAATA
AAAVGGGYKGNLSSSLAAVGGPPHPLSFPPFQDLNIYGNITDQVLDHNFHHQNIENH
FTGNNAADTNIYLGKKRPNPNFGNDVRKGLLMWSDQDHDLSANQSIDDEHRIQIQMATHV
STDLDLSLEIYERKSLSGDEGNNGKLLERPSPRRSPLSPMMPNGGLIQGRNSPFG*
>G727 (43..1977)
CTTCTTCTCTTCTCTGATCGTTCTGTTTCTGGACGAGAGAGATGGTAAATCCGGGTCAC
GGAAGAGGACCCGATTCCGGGTACTGCTGCTGGTGGGTCAAACCTCCGACCCGTTTCTGCG

AATCTTCGAGTTCTTGTCGTTGATGATGATCCAACCTTGCTCATGATCTTAGAGAGGATG
CTTATGACTTGTCTCTACAGAGAGCAGAGAGCGCATTGTCTCTGCTTCGGAAGAACAAG
AATGGTTTTGATATTGTCTATTAGTGATGTTTCATATGCCTGACATGGATGGTTTTCAAGCTC
CTTGAACACGTTGGTTTAGAGATGGATTTACCTGTTATCAATCTGAATGTTTTGAAACCT
TTGGTTATAGTGATGCTGCGGATGATTGGAAGAGCGTTGTGTTGAAAGGAGTGACTCAC
GGTGCAGTTGATTACCTCATCAAACCGGTACGTATTGAGGCTTTGAAGAATATATGGCAA
CATGTGGTGCGGAAGAAGCGTAACGAGTGGAATGTTTCTGAACATCTGAGGGAAGTATT
GAAGATACTGGCGGTGACAGGGACAGGCAGCAGCAGCATAGGGAGGATGCTGATAACAAC
TCGCTTTCAGTTAATGAAGGGAACGGGAGGAGCTCGAGGAAGCGGAAGGAAGAGGAAGTA
GATGATCAAGGGGATGATAAGGAAGACTCATCGAGTTTAAAGAAACCACGCGTGGTTTGG
TCTGTTGAATTGCATCAGCAGTTTGTGCTGCTGTGAATCAGCTAGGCGTTGACAGTGAG
TTAAAAACTTGCTTGTATGCAATTTGTGTGTGCTGATTGGTAACATTGTGGAATCCAG
AAGTATCGGATATATCTGAGACGGCTTGAGGAGGATATCGCAACACCAAGGAAATATGAAC
CATTGCTTTATGACTGGTCAAGATCAGAGTTTTGGACCTCTTCTTCGTTGAATGGATT
GATCTTCAATCTTTAGCTGTTACTGGTCAGCTCCCTCCTCAGAGCCTTGACACAGCTTCAA
GCAGCTGGTCTTGGCCGGCCTACACTCGCTAAACAGGGATGTCGGTTTCTCCCTTGTA
GATCAGAGAAGCATCTTCAACTTTGAAAACCCAAAAATAAGATTGGAGACGGACATGGT
CAGACGATGAACAATGGAAATTTGCTTCATGGTGTCCCAACGGGTAGTCACATGCGTCTG
CGTCTCGGACAGAAATGTTTCAGAGCAGCGGAATGATGTTGCCAGTAGCAGACCAGCTACCT
CGAGGAGGACCATCGATGCTACCATCCCTCGGGCAACAGCCGATATTGTCAAGCAGCGTT
TCAAGAAGAAGCGATCTCACTGGTGCCTGGCGGTTAGAAACAGTATCCCCGAGACCAAC
AGCAGAGTGTACCAACTACTCACTCGGTCTTCAATAACTTCCCCGCGGATCTACCTCGC
AGCAGCTTCCCGTTGGCAAGTGCCCCAGGGATTTCAGTTCCAGTATCAGTTTCTTACCAA
GAAGAGGTCAACAGCTCGGATGCAAAAGGAGGTTTCATCAGCTGCTACTGCTGGATTGGT
AACCAAGCTACGACATATTTAACGATTTTCCGAGCAGCAACAGCACAACAAGAATC
AGCAATAAACTAAACGATTTGGGATCTGCGGAATATGGGATTGGTCTTCACTTCCAATCAG
GACGACGCAACTGCAACCGCAACCGCAGCATTTTCCACTTCGGAAGCATACTCTTCGCT
TCTACGACAGAAAAAGACGGGAAACGGACGCAACAGTTGTGGGTGAGCATGGGCAGAAC
CTGAGTACCGAGCCGGAATCTGTATCATCTGAACACGTTTTTATGACGGTGGTTCA
GTCAGAGTGAAGTCAGAAAGAGTGGCGGAGACAGTGACTTGTCTCCAGCAAATACATTG
TTTACAGAGCAGTATAATCAAGAAGATCTGATGAGCGCATTTCTCAAACAGGTTTGATTA
TTACTCGAATACAGTGCCTCTAAAC

>G727 Amino Acid Sequence (domain in AA coordinates: 226-269)

MVNPGRHGRPDSTAGGSNSDPFANLRLVVDVDDPTCLMILERMLMTCLYREQRAHCL
CFGRTKNGFDIVISDVHMPDMDGFKLLEHVGLMDLPVNLNLKPLVIVMSADDSKSVV
LKGVTGHGAVDYLIKPVRIEALKNIWQHVVVRKKRNEWNVSEHSGSIEDTGDDRDRQQQHR
EDADNNSSSVNEGNRRSSRKRKEEEVDQDDKEDSSSLKKPRVVWSVELHQQFVAAVNQ
LGVDSELKTCCLMLHLCVIGNIVEFQKYRIYLRRLGGVSQHQGNMNHSMFTGQDQSFGL
SSLNGFDLQSLAVTGQLPPQSLAQLQAAGLGRPTLAKPGMSVSPVLDQRSIFNFENPKIR
FGDGHGQTMNNGNLLHGVPTGSHMLRPGQNVQSSGMLLPVADQLPRGGPSMLPSLGQQP
ILSSSVSRRLDGLAVRNSIPETNSRVLPTHSVFNNFPADLPRSSFPLASAPGISVP
VSVSYQEEVNSSDAKGGSSAATAGFGNPSYDIFNDFPQHQQHNKNISNKLNDWDLRNMGL
VFSSNQDAATATATAAFSTSEAYSSSTQQRKRETDATVVGEHGQNLQSPSRNLYHLNHV
FMDGGSVRVKSERVAETVTCPPANTLFHEQYNQEDLMSAFILKQV*

>G740 (25..924)

CTTCTTCACTTTTTTTTAAACGATGGCTTCAGAGGATCAATCGGCGGCGAGATCTACC
GGGAAGGTGAACCTGGTTCAACGCTTCTAAAGGCTATGGTTTCATTACTCTGACGATGGC
AGCGTAGAGCTTTTGTTCATCAATCTTCAATGTCTCCGAAGGTTACCGGAGTTTAACC
GTCGCGGATGCGGTTGAGTTGCTATTACTCAGGAAGCGACGGTAAGACTAAAGCCGTC
AATGTACTGCTCCTGGTGGTGGTTCTCTCAAGAAGGAGAATAACTCTCGTGGTAACGGT
GCTAGCGCGCGCGCGGTGGAAGCGGTTGCTACAATTGCGGTGAGTTAGGTTCATATCTCT
AAAGATTGTGTTATTGGTGGCGCGCGGAGGTGGTGAACGTAGATCTAGAGGAGGAGAA
GGTTGTTACAATTGTGGTGATCTGCTCACTTCGCTAGGATGTACTTCAGCTGGAAAC
GGTGAACAACCTGGAGCCACCAAGGTGGAACGATGGTTGCTACACTTGCCTGATGTT
GGTCACTGGCTAGGGATTGTACTCAGAAATCAGTTGGAACCGGAGACCAACGTGGAGCG
GTCAAAGGTGGAACGATGGTTGCTACACTTGTGGTGTGTTGGTCACTTGTCTAGGGAT
TGTAATCAGAAGGTTGCTGCCGGAACGTCAGAACCGGTGGTGGTGGTAGTGAACCTGT

TATTCATGCGGTGGAGTTGGTCACATTGCAAGAGATTGTGCGACTAAGAGACAGCCTTCT
CGTGGGTGTTACCAGTGTGGTGGTCTGGTCACTTGGCTCGTGATTGTGACCAGAGAGGA
AGCGGTGGAGGAGGTAATGATAATGCGTGCTACAAGTGTGGTAAGGAAGGTCACCTTTGCA
AGGGAATGTTCTTCTGTAGCTTAATCGATTTCCTAATCAACAAAACAAAAACAAGAAT
GAAATTGAATCGAGTTATATAGTTTGGTATATATACTCTTCGTTTTTCATTTATCTTTTT
TTTTGTTGTTGATGGGAATGAAATTGCCTGGTCCCTTTTGGTGTGTTTTTGAGCTTTTATT
ATTATACAGAGTGTATCCCTTTTTTGTATAACTATTACAAGTTTTTAGCTTTATTTGATA
TGGATGCTCTCTCTTTTCTTCTATCTGTTTCTGGAAATTTGACCTCATCATATTACTT
ATGTCATCCAAA

>G740 Amino Acid Sequence (domain in AA coordinates: 24-42, 232-268)

MASEDQSAARSTGKVNWFNASKGYGFITPDDGSVELFVHQSSIVSEGYRSLTVGDAVEFA
ITQSGDGKTKAVNVNTPAGGSLKKENNSRNGARRGGGSGCYNCGELGHISKDCGIGGG
GGGGERRSRGGEGCYNCGDTGHFARDCTSAAGNDQRGATKGGNDGCTCGDVGHVARDCT
QKSVNGDQRGAVKGGNDGCTCGDVGHFARDCTQKVAAGNVRSGGGSGTCYSCGGVGH
IARDCATKRQPSRGYQCGSGHLARDCDQRSGGGGNDNACYKCGKEGHFAECSSVA*

>G770 (119..1069)

CCTTCCTCTATATAAGGAAGTTCATTTTCATTTGGAGAGGACACGCTGACAAGCTGACTCT
AGCAGATCTGGTACCGTCGACGGTTCCTGGATTGGAGTAACTAAAGATCATATAAAAT
GGAAACAGGAGATCATCAGCAGCATAAGAAAGAAGAAGCTTTGCCACCGGGTTTCAG
ATTTTCATCCGACGGATGAGGAGCTAATCTCATATTACTTGGTTAATAAGATTGCCGATCA
AACTTCACCGGAAAGCAATCGCTGACGTTGATCTTAACAAGTCCGAGCCATGGGAGCT
TCCTGAGAAGCGGAAAATGGGAGGAAAAGAATGGTACTTTTTTAGCCTCCGGGACCGGAA
GTACCCGACCGGAGTGAGGACGAATAGGGCGACGAATACAGGATATTGGAAAACACAGG
AAAAGACAAAGAGATATTCAATAGCACAACTCGGAGTTGGTCCGGATGAAGAAGACTTT
GGTCTTTTACAGAGGACGAGCTCCTCGTGGGGAGAAGACTTGTGGGTCTGTCATGAGTA
TCGACTTCACTCCAAGTCTCATATAGAACCTCCAAGCAAGACGAGTGGGTAGTGTGTAG
AGTGTTCAGAAAACAGAAGCAACCAAGAAATACATAAGCACCAGTAGCAGCAGCACAAAG
TCATCACCACAACAACACACAAGAGCCTCAATACTATCAACCAACAACAATAATCCTAA
TACTCATCAGACCTCCTTCAACTCCCAACCGCATCTACAACACACCCGAGCCTCAATAT
TAACCAATCCCTCATGGCAAACGCCGTTACCTAGCTGAGCTCTCAAGAGTCTTCCGTGC
CTCTACAAGCACCACCATGGACTCTTCTCATCAGCAGCTAATGAACACACCCACATGCC
TGTCTCAGGGCTCAACCTCAACCTTGGCGGTGCACTGGTCCAGCCGCTCCTGTTGTGTC
TCTTGAGGATGTTGCGCGGTAGTGCTTCCGTACAATGGCGAAAACGGGTTTGAAATGT
GGAGATGAGCCAGTGCATGGACTTGGATGGATACTGGCCATCTTATTGATTGGTAATTGT
CAGTTTAAGTTATGTTTATATTGTTTCCATTTACTTGTGGTAAAACGATTTTGGTT
GTTCTTGCGAACCTCTAGACAGGCCTCGTACCGATCCTCTAGCTAGAGCTTTCGTTCTG
TATCATCGGTTTC

>G770 Amino Acid Sequence (domain in AA coordinates: 19-162)

MEQGDHQHKKKEEALPPGFRFHPTDEELISYYLVNKLADQNFQKAIADVLDNKSEPWE
LPEKAKMGKEWYFFSLRDRKYPTGVRTNRTNTGYWKTGKDKEIFNSTTSELVGMKKT
LVFYRGRAPRGEKTCWVMHEYRLHKSSTYRTSKQDEWVVCVFKTEATKKYISTSSSST
SHHHNNHTRASILSTNNNNPNYSSDLLQLPPLHQLPHSLNINQSLMANAVHLAELSRVFR
ASTSTTMDSSHQQLMNYTHMPVSGNLNLGGLVQPPPVVSLDVDAAVSASYNGENGFN
VEMSQCMDLDGYWPSY*

>G858 (99..869)

CATAATCTCTCTCTATATCTCTTCTCTTCTTTTACCCTGTTTTTTTTTTCATTC
CACAGAGCCAGGTTGATTGATTTTGTATTTCAGAGATATGGGAGAGGAAGGATTGAGA
TTAAGAAGATTGAGAATATCAACAGTCTGCAAGTCACTTTCTTAAGAGACGAAACGGTT
TGATCAAGAAGGCTAAAGAGCTTTGATTTCTGTGACGCCGAGGTTGCTCTTATCATCT
TCTCCAGCACCAGGCAAGATTTACGATTTCTCCAGCGTCTGTATGGAGCAAATCTTTCTA
GATATGGATACACTACTGCGTCCACTGAGCATAAACAACAAGAGAACACCAACTTCTAA
TTTGTGCTTCACATGGAAATGAAGCTGTGTTGCGAAATGATGATTCTGAAGGGGGAAC
TTGAAAGATTACAGCTTGCAATTGAGAGACTTAAGGGTAAGGAGCTTGAAGGTATGAGTT
TCCCGGATCTTATTTCTTGTGAAAACAGTTGAACGAGAGCTTGATAGTGTCAAGGATC
AAAAGACACAAATCTGCTCAACCAGATTGAGAGATCCAGGATACAGGAGAAAAAGCAT
TGAAGAAAAACCAATCTTGCGCAAACAGGTTGAGATGTTGGGAGAGGTTTCAGGACCA
AAGTGTGAATGAAAGGCCTCAAGATTCTAGCCCAGAAGCCGATCCCGAGAGCTTTCAT

CAGAAGAGGATGAGAATGACAACGAGGAGCACCATTCCGACACTTCCTTGCACTTGGGGT
TGTCGTCGACGGGGTATTGCACAAAGAGAAAGCCGAAGATCGAACTGGTCTGCGATA
ACTCTGGGAGTCAAGTGGCTTCTGATTGATGGAATCGATTATTTTCTAATTCTGGTTGT
TTAGGGGTCTCTATGTGTCTTCTGTTTCTGGCTGTTCTTTTGCTTTATTTTCATCTCAAG
TAGAGTTTCTTAATGTTTTAGGTGGAACATTTTCCATAATCAAGAAGGGATTGATCAA
TCAATAACATTAGATTTTCTTAGTTAAAGACTTAAAGTTGCCACACACCACACCATATG
TGATTATGATGAATTTACATTTTATAAAAAAAAAAAAAAAAAAAAAAAAAA
>G858 Amino Acid Sequence (domain in AA coordinates: 2-57)
MGRGRIEIKKIENINSRQVTFKRRNGLIKKAKELSILCDAEVALIIFSSTGKIYDFSSV
CMEQILSRYGYYTASTEHHKQREHQLLICASHGNEAVLRNDDSMKGELERLQLAIERLKG
KELEGMSFPDLISLENQLNESLHSVVDQKTQILNLQIERSRIQEKKALEENQILRKQVEM
LGRSGPKVLNERPDSSPEADPESSSSEEDENDNEEHSDTSLQLGLSSTGYCTKRKKP
KIELVCDNSGSQVAD*
>G865 (282..920)
ATCCCCACTTGTTGTTTCATCACCAAGCCAAGCTCCATGTCCTAGTCACTCCACAGATTCC
CTATCATCATCAATTCGTTTCAAACCTTAGTTCCTTTCAAAGTCTTGATACATATATACACA
CACACCTATTATTCTCTTGGTGTGTTTGTGTGTACATATACGTGTGAGTACATACTTTG
TTGTAAGTGGATCGGAGGTATGGAAGGGACCGGTTCCACCGGAAACATCGGCGCGCG
CGGATGATAATTCTGCTCTTGGAACGAGACTGATGTACCGCCATGGTCTCCGCTCTCAGCC
GTGTATAGAGAATCCGACAGACCCCGCGGTCAAACAAGAGCTTGATAAATCGGATCAAC
ATCAACCAGACCAAGATCAACCAAGAAGAAGACACTATAGAGGCGTAAGGCAGAGACCAT
GGGGTAAATGGGCGGCAGAAATCCGCGATCCAAAGAAAGCAGCCCGTGTCTGGCTCGGGA
CTTTTCGAGACGGCAGAGGAAGCTGCTTTAGCCTATGACCGAGCTGCCCTCAAATTCAAAG
GCACCAAGGCTAAACTGAACTTCCCTGAACGGGTCCAAGGCCCTACTACCACCACAACCA
TTTCTCATGCACCAAGAGGAGTTAGTGAATCCATGAACTCACCTCCTCCTCGACCTGGTC
CACCTTCAACTACTACTCTTCTGCGCAATGACTTATAACAGGACATACTTCAATACG
CTCAGTTGCTTACGAGTAACAATGAGGTTGATTATCATACTACACGTCGACTCTCTTCA
GTCAACCTTTTCAACGCCTTCTTCATCTTCTTCTTCTCCCAACAGACGCAGCAACAGC
AGCTACAACAACAACAACAGCAGCGTGAAGAAGAAGAGAATTATGGTTACAATTATT
ATAACTACCCAAGAGAATAATCTAATTATTATTGTTGGTTCGAATCAGTTTATAAATAGC
TATCATAGTTTTCATTTTGGTTTCCGTAACCTTTGTTGTCATGGAAAATATGAATGAACGA
GGGACATGTGTAACAATTTGTTTGTGTTTTCGTAAATGTTAGTTGATTGTTGGATTGCTGA
AGTTTGATTTTCTGAGCATAAATCATTTGACGGTCAAAAAAAAAA
>G865 Amino Acid Sequence (domain in AA coordinates: 36-103)
MVSALSRVIENPTDPPVKQLDKSDQHPDQDQPRRRHYRGVVRQRPWGKWAAEIRDPKKA
ARVWLGTFFETAEBAALAYDRAALKFKGTAKLNFPERVQGPTTTTTISHAPRGVSESMNS
PPPRPGPPSTTTTTPMTYNQDILQYAQLLTSNNEVDLSYTTSLFSQPFSTPSSSSSSS
QQTQQQQLQQQQQQREEEKNYGYNYNYNPRE*
>G872 (59..646)
CCGGAACAGAAATCAATTCAACCAAAACCGAATCGAACCGAACCGGAGTTTTTATCCAAT
GGTGAAGCAAGCGATGAAGGAAGAGGAGAAGAAGAGAAACACGGCGATGCAGTCAAAGTA
CAAAGGAGTGAGGAAGAGGAAATGGGGAAAATGGGTATCGGAGATCAGACTTCCACACAG
CAGAGAACGAATTTGGTTAGGCTCTTACGACACTCCCAGAAAGCGCGCGTGCTTTTCTGA
CGCCGCTCAATTTTGTCTCCGCGGCGCGGATGCTAATTTCAATTTCCCTAATAATCCACC
GTCGATCTCCGTAGAAAAGTCGTTGACGCCTCCGGAGATTGAGGAAGCTGCTGCTAGATT
CGCTAACACATTTCAAGACATTGTCAAGGGAGAAGAAGAATCGGGTTTAGTACCCGGATC
CGAGATCCGACCAGAGTCTCTTCTACATCTGCATCTGTTGCTACATCGACGGTGATTGA
TGATTTTTCGTTTTCGATTGCTTCCGATGAATTTTCGGGTTGATTCTTCTCCGACGA
CTTCTCTGGCTTCTCCGGTGGTGATCGATTTACAGAGATTTTACCCATCGAAGATTACGG
AGGAGAGAGTTTATTAGATGAATCTTTGATTCTTTGGGATTTTGAATTCCCAAACATAA
TATTTTCTTAGAGCGAAGTGTGAGATTTTCTTGGAGTCATGGAGAAATCTGGAGATTTT
TTGTAACACGGAGCTCCAATGACCCGGGAATTTCTTCTGTTTCGGATCCGAATTTGATGT
GGATCATATTACACCTATATTTTTCATTTTTTGTGTAAGAAAAATCGGATAAGAT
TCTAGTAATAAATGTAAAGTCCATTTTCAAAAAAAAAAAAAAAAAAAAAA
>G872 Amino Acid Sequence (domain in AA coordinates: 18-85)
MVKQAMKEEEKRNTAMQSKYKGVKRKWKVSEIRLPHSRERIWLGSYDTPEKAARAF
DAAQFCLRGGDANFNFPNPPSISVEKSLTPPEIQEAAARFANTFQDIVKGEEESGLVPG

SEIRPESPSTASVATSTVDYDFSFLDLLPMNFGFDSFSDDFSGFSGGDRFTEILPIEDY
GGESLLDESLILWDF*

>G904 (1..1005)

atggaatctctcatcaatcccagccatggcggaggaaactacgattctcactcttcttct
ctcgatagctctcaaaccaagcgtagtcatcattctcattctcctcatgactcttctc
atctccggtttccatttgcttctcctcctcgctgtctcaatcgctgtagccaccgctccggt
ctcctcttctcatcttctccttccgctcgcaaccgtaacttccgattcccgcagattctct
ggacatcgagtctctccgaacagaacggctcctcgtgcttgattcgcttccgattttc
aaattctcctccgctcactcgccgatctagctccatgaattccggagattgcgccggttgt
ttgtcgaaattcgaaccggaggatcagctccgctcttctcctctctgttgacgctttt
cacgcccattgtatcgatatctggctagtctctaaccagacttgctctctctgctgctct
cctctcttctgcttcagaatctgatctcatgaagtctctcgccgtcgctcggtcaaacaac
ggcggaggagaaaacagcttccgctctcgaatcggatccatcagccgctcgctcgaaca
ccgattccagaatccggttgagcagcatcgaacttactcaatcggttgcgttcgattacata
gtagacgacgtagattcagaatctcagagtcaaatttcaaccgtggaaaacaggaagac
gcgactacaacaactgccacagcaacggcggttacgactaatccgacgtcggttgagct
agtttagcggcggtataggtaacgatggttctagaagctggctcaaggattacggttgac
agactctcacgaggtatctcgctcgctgcaatgtcggttagaagctctggttagattttt
actgggagtagtctcgaggtgaggaattgacggtgatggatttagaagcgaatcatgcc
ggagaagagataagtgagcttttccggtggctctcaggggtgtga

>G904 Amino Acid Sequence (domain in AA coordinates: 117-158)

MESLINPSHGGGNYDSSHSSSLSLKPSVLVIIILILLMTLLISVSICFLRLNRCSHRSV
LPLSSSSSVATVTSDSRFRSGHRVSPETERSSVLDLPIFKFSSVTRRSSMNSGDCAVC
LSKFEPEDQLRLPLCCHAFHADCIDIWLVSNQTCPLCRSPLFASESMLKSLAVVGSNN
GGGENSFRLEIGSISRRTPIPIESVEQHRYSIGSFDYIVDDVDSEISESNFNRGKQED
ATTTTATATAVTNTPTSFASLAADIGNDGRSRLKDYVDRLSRGISSRAMSFRSSGRFF
TGSSRRSEELTVMDLEANHAGEEISELFRWLSGV*

>G910 (1..1071)

ATGTTATGTATAATAAATGAGAATATGGAAAGAGTATGTGAGTTTGTAAAGCGTAT
AGAGCAGTGGTTTATTGTATAGCTGATACAGCAAATCTTTGTTTAAATGTGATGCAAAG
GTTCAATTCAGCTAATTCACCTCTCGGGACGGCATTACGTACGGTTTATGTGATTCTGGT
AAGAATCAGCCTTGTGTTGTCGGATGTTTGGACCATAAAATGTTCTTTGCCATGGATGT
AATGATAAGTTTCATGGTGGTGGCTCTTCTGAGCATCGTAGAAGGGATTGAGGTGTTAT
ACGGGTTGTCTCTCTGCTAAAGATTTCCGGTTATGTGGGGTTTTCGAGTTATGGATGAC
GATGATGATGTTTCGTTAGAGCAATCTTTTGAATGGTAAACCTAAGGTGCAAAGAGAA
GGTGGTTTTATCTTGAACAGATTCTTGAATTGGAGAAGGTTACGCTCAGGGAAGAGAAT
GGTAGTTCTTCTTGACAGAACGAGGTGATCCATCTCCATTGGAGCTTCTTAAGAAACCC
GAAGAACAGTTAATCGATCTTCCGCAGACCGGAAAGAGCTGGTTGTTGATTTTTCACAC
TTGTCCTCATCTTCCACTTGGTGATTCCTTTTGGGAATGCAAAGTCCATACAATAAG
AACAATCAGTTGTGGCATCAAAATATACAAGACATTGGAGTATGTGAAGATACAATCTGC
AGTGACGATGACTTCCAAATACCTGACATTGATCTCACTTTCCGGAACCTTGAAGAGCAA
TTTGGAGCTGATCCTGAGCCAATTGCAGATAGTAACAACGTTCTTTGTTTCTTCCCTT
GACAAATCACATGAGATGAAGACATTTCTTCTTCAATCAATAATCCCATATTTGCACCT
AAACGAGCTTCATCACTATCTCATCTCAAGCAGTGAAACCGATAACCCCTTATAGTCAC
TCAGAGGAAGTAATCTCATTTTGTCCCTCCCTCTCTAACAATACACGTCAAAGGTCATC
ACAAGGCTCAAGGAGAAGAAGAGCAAGAGTGGAGGAGAAAAAGCTTAA

>G910 Amino Acid Sequence (domain in AA coordinates: 14-37, 77-103)

MLCIIIIENMERVCFCKAYRAVVYCIADTANLCLTCDKAVHSANSLSGRHLRTVLCDSG
KNQPCVVRFCDFHKMFLCHGCNDKFHGGSSSEHRRDLRCYTGCPKADFAVMWGFVMD
DDVSLEQSFMRMVKPVQREGGFILEQILELEKVLREENGSSSLTERGDPSPLELPKPP
EEQLIDLPLQTKELVVDFSHLSSSSTLGDSFWECKSPYNKNNQLWHQNIQDIGVEDTIC
SDDDFQIPDIDLTFRNFEEQFADPEPIADSNVFFVSSLDKSHMKTFSSSFNNPIFAP
KPASSTISFSSSETDNPYSHSEEVISFCPSLSNNTRQKVITRLKEKKRARVEEKKA*

>G912 (20..694)

CATCTTATCCAAAGAAAAAATGAATCCATTTTACTCTACATTCCCAGACTCGTTTCTCTC
AATCTCCGATCATAGATCTCCGGTTTCAGACAGTAGTAGTGTTCACCAAAGTTAGCTTC
AAGTTGTCCAAAGAACGAGCTGGGAGGAAGAAGTTTCGTGAGACACGTCATCCGATTTA

CAGAGGAGTTCGTCAGAGGAATTCTGGTAAATGGGTTTGTGAAGTTAGAGAGCCTAATAA
GAAATCTAGGATTTGGTTAGGTACTTTTCCGACGGTTGAAATGGCTGCTCGTGCTCATGA
TGTTGCTGCTTTAGCTCTTCTGGTGGTCTGCTTGTCTCAATTTTCGCTGATTCTGCTTG
GCGGCTTCGTATTCCTGAGACTACTTGTCTTAAGGAGATTAGAAAGCTGCGTCTGAAGC
TGCAATGGCGTTTCAAGATGAGACTACGACGGAGGGATCTAAACTGCGGCGGAGGCAGA
GGAGGCGGAGGGGAGGGGGTGAAGGAGGGGAGAGGAGGGCGGAGGAGCAGAATGGTGG
TGTGTTTTATATGGATGATGAGGCGCTTTTGGGGATGCCCAACTTTTTTGAGAATATGGC
GGAGGGGATGCTTTTGCCGCCGCCGGAAGTTGGCTGGAATCATAACGACTTTGACGGAGT
GGGTGACGTGTCACTCTGGAGTTTGTACGAGTAATTTTTTGGCTCTTTTTCTGGATAATA
AGTT

>G912 Amino Acid Sequence (domain in AA coordinates:51-118)
MNPFFYSTFPDSFLSISDHRSPVSDSSECPKLASSCPKKRAGRKKFRETRHPIYRGVRQR
NSGKWVCEVREPNNKSRWLGTFTVEMAARAHDAALALRGRSACLNFAWSAWRLRIPE
TTCPEIQKAASEAAMAFQNETTEGSKTAAEAEEAAGEGVREGERRAEEQNGGVFYMD
EALLGMNFFENMAEGMLLPPEVGNHNDFFDGVDSLSWSFDE*

>G920 (114..1154)
AAAAATCTATTTCTTCTCTTCCACTATATTACAACATTTCTTCATTCTCAAATCATC
ATACTAAAAACCTAAAAAAGTTACATATTCAATTGTATCTTTGTGAGAAAAAATGGATT
CGAATAGTAACAACACGAAATCCATAAAGAGAAAAAGTTGTGACCAACTTGTGGAAGGCT
ATGAATTCGCTACTCAGCTTCAGCTTCTCCTTTCTCATCAACACTCTAACCAGTACCACA
TCGATGAGACCCGCTTGTGTTCCGGGTTCGGGTTTCAGTTTCCGGTGGTCCAGATCCCGTTG
ATGAGCTCATGTCTAAGATCTTGGGATCTTTCCATAAAACTATATCGGTTCTTGATTCTT
TTGATCCCGTCGCGCTCTCTGTCCCATCGCCGTCGAGGGTTTCATGGAATGCTTCATGTG
GGGATGATTCCGCGACTCCGGTGAGTTGCAACGGTGGAGATTCCGGTGAGAGTAAGAAGA
AGAGATTAGGGGTTGTTAAGGAGGTAAGAGGATGCTACACTAGAAAGACGAGATCACATA
CAAGGATCGTGGAAGCTAAAGTTCTGAAGACAGATATGCTTGAGGAAATATGGACAAA
AGGAGATTCTTAATACCACATTCCCAAGAAGTTACTTTAGATGCACACACAAGCCAAACGC
AAGGATGCAAAGCAACAAGCAAGTTCAAGAACAGGATCAAGATTCTGAGATGTTCCAAA
TCACATACATTGGCTACCACACATGCACTGCCAATGACCAACGCACGCGAAGACCGAGC
CTTTTGATCAAGAAATCATTATGGATTCCGAAAAGACATTGGCTGCTAGCACTGCTCAGA
ACCATGTCAATGCTATGGTGCAAGAGCAAGAGAACACACCAGCAGTGTGACAGCAATAG
ACGCAGGCTAGGGTTGTTAAGGAGGTAAGAAATAACAATGGTGATCAGAGTAAGATTATTATG
AGGGCTCTTCGACAGGTGAGGACTTGTCAATTGGTTTGGCAAGAGACGATGATGTTTGATG
ATCATCAAAATCACTACTATTGTGGTGAAACCAGTACTACTTCTCATCAATTTGGTTTCA
TCGACAACGATGATCAGTTTTCCTCCTTCTTCGACTCATATTGTGCTGATTATGAAAGAA
CAAGTGCTATGTGAACATCCAAATCTGGAATGATGAATCAGCACTAGGTCTTCTCTTTGA
GTATGTCTAGTTTAATGTAATATTTTGTGTATGTTTGATAAACAACCATATATACTT
CTCTTTTACACCAAAAAAAAAAAAAAAAAAAAAA

>G920 Amino Acid Sequence (domain in AA coordinates: 152-211)
MDSNSNNTKSIKRKVVQDLVEGYEFATQLQLLLSHQHSNQYHIDETRLVSGSGSVSGGPD
PVDELMSKILGSFHKTI SVLDSFDPVAVSVPIAVEGSWNASCDDSATPVSCNNGDSGES
KKKRLGVGKGRGCYTRKTRSHTRIVEAKSSEDRYAWRKYQKEILNTTFPRSYFRCTHK
PTQGCATKQVQKQDQDSEMFQITYIGYHTCTANDQTHAKTEPFQDEIIMDSEKTLAAT
AQNHVNAMVQEQENNTSSVTAIDAGMVKEEQNNNGDQSKDYEGSSSTGEDLSLVWQETMM
FDDHQNHYYCGETSTTSHQFGFIDNDQFSSFFDSYCADYERTSAM*

>G939 (9..1565)
CAGATTCTATGGATATGTATAACAACAATATAGGGATGTTCCGGAGTTTAGTTTGTAGCT
CGGCGCCTCCATTTACAGAGGGACATATGTGTTCTGATTGCGATACGGCTTTGTGCGATG
ATCTGAGTAGTGATGAGGAAATGGAATAGAGGAGCTTGAGAAGAAGATCTGGAGAGACA
AGCAGCGTTTAAAGCGGCTCAAGGAAATGCGGAAGAACGGTCTAGGAACAAGATTGTTGT
TGAAGCAGCAACATGATGATTTTCCAGAGCACTCTAGTAAGAGAACCATGTACAAGGCAC
AAGATGGGATCTTGAAGTACATGTCGAAGACAATGGAGCGATATAAAGCTCAAGGTTTGT
TTTATGGGATTGTGTTAGAGAATGGGAAAACGGTAGCGGGATCTTCTGATAATCTCCGTG
AATGGTGGAAGACAAAGTGAGGTTTGTATAGGAACGGCCAGCTGCTATAATCAAGCACC
AAAGGGATATCAATCTTCTGATGGAAGTGATTCAAGGCTGAGGTTGGGGATTCTACCG
CACAGAAGTTGCTTGAGCTTCAAGATACTACTCTTGGAGCTCTGTTATCGGCTCTGTTTC
CTCACTGCAACCTCTCAGAGGCGGTTTCCGTGGAGAAAGGCGTGACACCGCCATGTT

GGCCAACGGGGAAAGAAGATTGGTGGGATCAACTGTCTTTACCCGTTGATTTTCGAGGTG
TTCCGCCACCTTACAAGAAGCCTCATGATCTCAAGAAGCTGTGGAAAATTGGTGTCTTTGA
TTGGTGTAATCAGACATATGGCTTCTGACATTAGCAACATACCCAATCTCGTGAGACGGT
CTAGAAGTTTGCAGGAGAAAATGACGTCAAGAGAAGGCGCTTTATGGCTCGCTGCTCTTT
ACCGAGAAAAGGCTATTGTTGATCAAATAGCCATGTCTAGAGAAAACAACAACACTTCTA
ACTTTCTTGTTCCTGCAACCGGTGGAGACCCAGATGTTTTGTTTCTGAATCTACAGACT
ATGATGTTGAACTGATTGGTGGCACTCATCGGACCAATCAGCAGTATCCTGAATTTGAAA
ACAACACTCACTGTGTTTACAAGAGAAAAGTTTGAAGAAGATTTTGGGATGCCAATGCATC
CAACACTCCTAATCATGTGAGAACAGTCTCTGTCTTATAGCCAACCACATATGGGATTTT
TTGACAGGAACCTAAGAGAGAATCACCAAATGACTTGTCTTATAAAGTCACTTCTCTCT
ACCAACCAACTAAACCTTATGGTATGACGGGTTTAAATGGTTCCTTGTCCGGATTATAACG
GGATGCAGCAGCAGGTTTCAAGAGCTTCAAGACCACTTAAATCATCCCAACGATCTCTACA
GACCAAAAGCTCCACAAAGAGGCAACGATGACTTGGTTGAGGATTTGAATCCTTCTCTCT
CGACGCTGAATCAGAACTTGGTTTAGTCTTACCTACTGACTTCAATGGAGGTGAGGAAA
CAGTAGGAACAGAGAACAAATCTGCATAATCAAGGGCAAGAGTTGCCACATCTTGGATTCT
AGTAAAGAAAGCTTCAAGAGTTTCTTTTATGTTTCTAGTCTTTATAGCTTGTCTCTT
GCTTATCTCTCATTAAACACAGTTTTGTATCTCTCCATTTTCATAGCCCATGTAGCAATG
GAGAAGATTAGGTTTATAATAAGTTAATAACCAAATTCAA
>G939 Amino Acid Sequence (domain in AA coordinates: 97-106)
MDMYNNNIGMFRSLVCCSSAPPFTEGHMCSDSHTALCDDLSSDEEMIEIELEKKIWRDKQR
LKRLEKEMAKNGLGTRLLKQHQDDFPEHSSKRTMYKAQDGILKYMSKTMERYKAQGFVYG
IVLENGKTVAGSSDNLEWWDKVRFDNRGPAALIKHQRDINLSDGSDSGSEVGDSTAQK
LLELQDITLGLALLSALFPHCNPPQRRFPLEKGVTPPWPTGKEDWWDQLSLPVDVFRGVPP
PYKKPHDLKKLWKIGVLIGVIRHMASDISNIPNLVRRSRLQEKMTSREGALWLAALYRE
KAIVDQIAMSRENNNTSNFLVPATGGDPDLFPPESTDYDVELIGGTHRTNQOYPEFENNY
NCVYKRKFEEDFGMPMHPTLLTCENSLCPYSQPHMGFLDRNLRENHQMTCPIKVTSPYQP
TKPYGMTGLMVPCPDYNGMQQVQSFQDQFNHPNDLYRPAKQGRNDDLVEDLNPSPSTL
NQLGLVLPDFTDFNGGEETVGTENNHLNQQLPTSWIQ*
>G963 (1..897)
ATGAGTTTGCCTCCAGGATTCAAGTTTCACTCCCATGATGAAGAAGTGGTGGCTTACTAT
CTTGATAGGAAGGTCAACGCCAAGCCATTGAGCTCGAGATCATCCAGAAGTTGATCTT
TATAAATGCGAGCCATGGGACTTGCCTGAAAAGTCACTTTTTCGCCGGAAACGACATGGAA
TGGTACTTTTACAGCACAGGGATAAGAAGTATCCAAATGGCTCTAGGACGAACCGTGCG
ACCGGAGCGGTTACTGGAAGGCCACGGGGAAAGATCGTACAGTAGAATCAAAGAAGATG
AAGATGGGAATGAAGAAGACACTGGTTTATTATAGAGGAAGGGCTCCTCATGGCCTTCGT
ACTAATTGGGTATGCATGAATATCGTCTCACGCACGCTCCTTCTCTCTTGAAGGAG
TCGTATGCATTTGTGCCGAGTGTGTTAAGAAGAACATACAAATCCAAAGAGAAAAGGGGAA
GAAGAAGAAGCAGAAGAGAGAGCACTAGTGTAGGAAAAGAAGAGGAAGAAGAAAAGGAG
AAGAAGTGGAGAAAATGTGATGGTAATTATATTGAAGACGAGAGCTTGAAAAGAGCATCC
GCGGAGACATCTTCATCAGAGCTAACTCAAGGGGTCCTTTTAGACGAAGCAAAACAGCTCA
TCCATATTTGCTCTTCTATTTCTCTCTCTCTCTGACGATCATGATCATCTTTTCTCA
AACTATTCTCATCAGCTTCCATATCATCTCTCTCTCACTCCAAGATTCCCTCAACTT
TCTATGAACGAAGCAGAGATTATGTCAATCCAACAAGACTTTCAATGCAGAGACTCTATG
AACGGGACACTTGACGAAATCTTCTCTTCTCCGCCACTTTCCCGCTTCCCTTTGA
>G963 Amino Acid Sequence (domain in AA coordinates: TBD)
MSLPPGFRFHPTDEBLVAYYLDKRVNGQAELEIIEVDLYKCEPWLPEKSFLPGNDME
WYFYSTRDKKYPNGSRNTRATAGYWKATGKDRTVESKKMKMGKTLVYYRGRAPHGLR
TNWVMHEYRLTHAPSSSLKESYALCRVFKKNIQIPKRKGEEEEEESTSVGKEEEEEKE
KKWRKCDGNYIEDESLKRASAEETSSSELQGVLLDEANSSSIFALHFSSLLDDHDLFS
NYSHQLPYHPPLQLQDFPQLSMNEAEIMSIQQDFQCRDSMNGTLDEIFSSSATFPASL*
>G979 (60..1352)
CCTCTGAGGAATCAAATCACTCACACTCCAAAAAATCTAACTTTCTCAGAGTTTAA
TGAAGAAGCGCTTAACCACTTCCACTTGTCTTCTTCTCCATCTTCTCTGTTTCTTCTT
CTACTACTACTTCTCTCTTCTCAGTCGGAGGCTCCAAGGCCTAAACGAGCCAAAAGGG
CTAAGAAATCTTCTCTTCTGGTGATAAATCTCATAACCCGACAAGCCCTGCTTACCC
GACGCAGCTCTATCTACAGAGGAGTCACTAGACATAGATGGACTGGGAGATTGAGGCTC
ATCTTTGGGACAAAAGCTCTTGAATTTCGATTGAGAACAAGAAAGGCAACAAGTTTATC

TGGGAGCATATGACAGTGAAGAAGCAGCAGCACATACGTACGATCTGGCTGCTCTCAAGT
ACTGGGGACCCGACACCATCTTGAATTTTCCGGCAGAGACGTACACAAAGGAATTGGAAG
AAATGCAGAGAGTGACAAAGGAAGAATATTTGGCTTCTCTCCGCCAGAGCAGTGGTT
TCTCCAGAGGCGTCTCTAAATATCGCGGCGTCGCTAGGCATCACCACAACGGAAGATGGG
AGGCTCGGATCGGAAGAGTGTGTTGGGAACAAGTACTTGTACCTCGGCACCTATAATACGC
AGGAGGAAGCTGCTGCAGCATATGACATGGCTGCGATTGAGTATCGAGGCGCAAACGCGG
TTACTAATTTTCGACATTAGTAATTACATTGACCGGTTAAAGAAGAAAGGTGTTTTCCCGT
TCCCTGTGAACCAAGCTAACCATCAAGAGGGTATTCTTGTGTAAGCCAAACAAGAAGTTG
AAACGAGAGAAGCGAAGGAAGAGCCTAGAGAAGAAGTGAAACAACAGTACGTGGAAGAAC
CACCGCAAGAAGAAGAAGAGAAGGAAGAAGAGAAGCAGAGCAACAAGAAGCAGAGATTG
TAGGATATTGAGAAGAAGCAGCAGTGGTCAATTGCTGCATAGACTCTTCAACCATAATGG
AAATGGATCGTTGTGGGGACAACAATGAGCTGGCTTGGAACTTCTGTATGATGGATACAG
GGTTTTCTCCGTTTTTGGCTGATCAGAATCTCGCGAATGAGAATCCCATAGAGTATCCGG
AGCTATTCAATGAGTTAGCATTTGAGGACAACATCGACTTCATGTTTCGATGATGGGAAGC
ACGAGTGTGTAACCTTGGAAAATCTGGATGTTGCGTGGTGGGAAGAGAGAGCCCACCCCT
CTTCTTCTTACCATTGTCTTGCTTATCTACTGACTCTGCTTCATCAACAACAACAACAA
CAACCTCGGTTTTCTGTAACTATTTGGTCTGAGAGAGAGAGCTTTGCCTTCTAGTTTGAA
TTTCTATTTCTCCGCTTCTTCTTTTCTTTTCTTTTGTGGGTTCTGCTTAGGGTTTG
TATTTAGTTTCCGGCTTGTTCGTTGGTCTGAATAATCAATGTCTTTGCCCTTTTNN
AANGNTNCAAGNTNAAANAAAAA

>G979 Amino Acid Sequence (domain in AA coordinates: 63-139,165-233)

MKRLTSTSCSSPSSSVSSSTTTSSPIQSEAPRPKRAKRAKSSPSGDKSHNPTSPAST
RRSSIYRGVTRHRWTGRFEHLWDKSSWSN IQNKKGKQVYLGAYDSEEAHAHYDLAALK
YWGPDITLNFPAETTYKELEEMQVRVTKEEYLASLRQSSGFSRGVSKYRGVARHHHNGRW
EARIGRVFGNKYLYLGTINTQEBAAAAYDMAAIEYRGANAVTNFDISNYIDRLKKKGVP
FPVNQANHQEGILVEAKQEVETREAKEEPREEVKQYVEEPPQEEEEKEEKEAEQQAEEI
VGYSEEAADVNCIDSSIMEMDRCDGNNELAWNFCMMDTGFSPLTDQNLANENPIEYP
ELFENELAFEDNIDFMFDDGKHECLNLENLDCCVVGRESPPSSSSPLSCLSTDSASSTTTT
TTSVSCNYLV*

>G987 (1..4011)

ATGGGTTCTTACTCAGCTGGCTTCCCTGGATCCTTGGACTGGTTTGATTTTCCCGGTTTA
GGAAACGGATCCTATCTAAATGATCAACCTTGTAGATATTGGATCTGTTCTCTCCTCCT
CTAGACCCATATCTCAACAGAACTTGGCTTCTGCGGATGCTGATTTCTCTGATTCTGTT
TTGAAGTACATAAGCCAAGTCTTATGGAAGAGGACATGGAAGATAAGCCTTGTATGTTT
CATGATGCTTTATCTCTTCAAGCAGCTGAGAAGTCTCTCTATGAAGCTCTCGGCGAGAAG
TACCGGTTTGATGATTCTGATCAGCCTCTGACTACTACTAGCCTTGCTCAATTGGTT
AGTAGTCCTGGTGGTTCTTCTTATGCTTCAAGCACCACAACCACTTCTCTGATTCAAA
TGGAGTTTTGATTGTTTGGAGAATAATAGGCCTTCTTCTTGGTTGCAGACACCGATCCCG
AGTAACTTCATTTTTCAGTCTACATCTACTAGAGCCAGTAGCGGTAACGCGGTTTTCGGG
TCAAGTTTTAGCGGTGATTTGGTTTCTAATATGTTAATGATACTGACTTGGCGTTACAA
TTCAAGAAAGGGATGGAGGAAGCTAGTAAATTCCTTCTAAGAGCTCTCAGTTGGTTATA
GATAACTCTGTTCTTAACAGATTAACCGAAAGAAGAGCCATTGGCGCGAAGAAGACAT
TTGACTGAAGAAAGAAGTAAGAAACAATCTGCTATTTATGTTGATGAAACTGATGAGCTT
ACTGATATGTTTGACAATATTCTGATATTTGGCGAGGCTAAGGAACAACCTGTATGCATT
CTTAACGAGAGTTTCCCTAAGGAACCTGCGAAAGCTTCAACGTTTAGTAAGAGTCCTAAA
GGCGAAAACCGGAAGCTAGTGGTAACAGTTATACAAAAGAGACACCTGATTTGAGGACA
ATGCTGGTTTTCTTGTGCTCAAGCTGTTTTCGATTAACGATCGTAGAACTGCTGACGAGCTG
TTAAGTCGGATAAGSCAACATTCTTCATCTTACGGCGATGGAACAGAGAGATTGGCTCAT
TATTTTGCTAACAGTCTTGAAGCACGTTTGGCTGGGATAGGTACACAGTTTTTACTGCC
TTGTCTTCCAAGAAACATCTACTTCTGACATGTTGAAAGCTTATCAGACATATATATCA
GTCTGTCCGTTTCAAGAAAATCGCAATCATATTCGCCAACCATAGTATTATGCGGTTGGCT
TCAAGTGCTAATGCCAAAACCATCCACATCATAGATTTTGGAAATATCTGATGGTTTCCAG
TGGCCTTCTCTGATTCATCGACTTGGCTTGGAGACGTGGTTTCTGTAAGCTTCCGATA
ACCGGTATAGAGTTGCTCAACGCTGTTTGTAGACAGCGGAGGAGTTATTGAGACTGGT
CGTCGCTTGGCTAAGTATTGTGAGAAGTTCAATATTCCGTTTGGAGTACAATGCGATTGCG
CAGAAATGGGAATCAATCAAGTTGGAGGACTTGAAGCTAAAAGAAGGCGAGTTTGTGCG
GTAAACTCTTTATTTCCGGTTTAGGAATCTTCTAGATGAGACGGTGGCAGTGCATAGCCCCG

AGAGATACGGTTTTGAAGCTGATAAGGAAGATAAAGCCAGACGTGTTTCATCCCCGGGATC
CTCAGCGGATCCTACACGCGCCTTTCTTTGTACAGAGTTTAGAGAAGTTCTGTTTCAT
TACTCATCTCTGTTTGACATGTGTGACACGAATCTAACCGGGAAGATCCAATGAGGGTT
ATGTTTGAGAAAGAGTTCTATGGGCGGGAGATCATGAACGTGGTGGCGTGTGAGGGGACG
GAGAGAGTGGAGAGGCCAGAGAGTTATAAGCAGTGGCAGGCGAGGGCGATGAGAGCCGGG
TTTAGACAGATTCGGCTGGAGAAGGAAGTCTAGTTCAGAACTGAAGTTGATGGTGGAAAGT
GGATACAAACCCAAAGAGTTTGATGTTGATCAAGATTGTCACTGGTTGCTTCAGGGCTGG
AAAGGTAGAATTGTATACGGTTCATCTATTTGGGTTCCCTTTCTTTTCTATGTGGGCAGA
GCAACTAGGGTTTGATCATGGATCCAACTTCTCTGAATCTCTAAACGGCTTTGAGTAT
TTTGATGGTAACCTAATTTGCTTACTGATCCAATGGAAGATCAGTATCCACCACCATCT
GATACTCTGTTGAAATACGTGAGTGAGATTCTTATGGAAGAGAGTAATGGAGATTATAAG
CAATCTATGTTCTATGATTCAATTGGCTTTACGAAAACTGAAGAAATGTTGCAGCAAGTC
ATTACTGATTCTCAAAATCAGTCCCTTTAGTCTCTGCTGATTCAATTGATTACTAATCTTGG
GATGCAAGCGGAAGCATCGATGAATCGGCTTATTCGGCTGATCCGCAACCTGTGAATGAA
ATTATGGTTAAGAGTATGTTTAGTGATGCAGAATCAGCTTTACAGTTTAAGAAAGGGGTT
GAAGAAGCTAGTAAATTCCCTTCCCAATAGTGATCAATGGGTTATCAATCTGGATATCGAG
AGATCCGAAAGGCGGATTCGGTTAAAGAAGAGATGGGATTGGATCAGTTGAGAGTTAAG
AAGATCATGAAAGGGATTGAGGAAGTTAGGAGTAGTAAGCAATTTGCTAGTAATGTA
GAAGATAGTAAGGTTACAGATATGTTTGATAAGGTTTGGCTTCTTGACGGTGAATGCGAT
CCGCAACATTGTTAGACAGCGAGATTCAAGCGATTCCGAGTAGTAAGAACATAGGAGAG
AAAGGGAAGAAGAAGAAGAAGAAGAGTCAAGTGGTGTATTTTCGTACACTTCTCACT
CATTGTGCACAAGCCATTCCACAGGAGATAAAACCACGGCTCTTGAGTTTCTGTTACAG
ATAAGGCAACAGTCTTCGCTCTCGGTGACGCGGGGCAAGACTAGCTCATTGTTTCGCT
AACCGCTTGAAGCTCGTCTACAGGGAAGTACCGGCTCTATGATCCAGACTTATTACAAT
GCTTTAACCTCGTCTGTTGAAGGATACTGCTGCGGATACAATTAGAGCGTATCGAGTTTAT
CTTTCTTCGCTCCGTTTGTACCTTGATGATTTCTTCTCCATCTGGATGATTCTTGAT
GTGGCTAAAGATGCTCCTGTTCTTATATAGTTGATTTTGGGATTCTATACGGGTTTCAA
TGGCCGATGTTTATTAGTCTATATCAGATCGAAAAGATGTACCGCGGAGCTGCGGATT
ACTGGTATCGAGCTTCCCTCAGTGCAGGTTTTCGGCCCGGAGCGAATAGAGGAGACAGGA
CGGAGATTGGCTGAGTATTGTAACCGGTTTAAATGTTCCGTTTGAGTACAAAGCCATTGCG
TCTCAGAACTGGGAAACATCCGGATAGAAGATCTCGATATACGACCAACGAAGTCTTA
GCGGTTAATGCTGGACTTAGACTCAAGAACCCTCAAGATGAAACAGGAAGCGAAGAGAAT
TGCCCCGAGAGATGCTGTCTTGAAGCTAATAAGAAACATGAACCCGGAGCTTTTCATCCAC
GCGATTGTTCAACGGTTCATTCAACGCACCTTCTTTATCTCGCGGTTTAAAGAAGCGGTT
TACCATTACTCCGCTCTCTTCGACATGTTTGATTCGACGTTGCCTCGGGATAACAAAGAG
AGGATTAGGTTTCGAGAGGGAGTTTACGGGAGAGAGGCTATGAACGTGATAGCGTGCGAG
GAAGCTGATCGAGTGGAGAGGCTGAGACTTACAGGCAATGGCAGGTTAGAATGGTTAGA
GCCGGGTTTAAAGCAGAAAACGATTAAGCCTGAGCTGGTAGAGTTGTTTAGAGGAAAGCTG
AAGAAATGGCGTTACCATAAAGACTTTGTGGTTGATGAAAATAGTAAATGGTTGTTACAA
GGCTGGAAAGGTCGAACCTCTCTATGCTTCTTCTTGTGGGTTCCCTGCCTAG
>G987 Amino Acid Sequence (domain in AA coordinates: 428-432,704-708)
MGSYSAGFPGLDWFDFPGLGNGSYLNDQPLLDIGSVPPPLDPYPQONLASADADFSDSV
LKYISQVLMEDMEDKPCMFHDALSLQAAEKSLYEALGEKYPVDDSDOPLTTTTSLAQLV
SSPGGSSYASSTTTSSDSQWFSFDCLENNRPSSWLQTPIPSNFIFQSTSTRASSGNAVFG
SSFSGDLVSNMFNDTDLALQFKGMEEASKFLPKSSQLVIDNSVNPRLTGKKSHWREEEH
LTEERSKKQSAIYVDETDELTFMFDNILIFGEAKEQPVCIILNESFPKEPAKASTFSKSPK
GEKPEASGNSYTKETPDRLTMLVSCAQAVSINDRRTADELLSRIRQHSSSYGDTERLAH
YFANSLEARLAGIGTQVYTTALSSKKTSTSDMLKAYQTYISVCPFKKIAIIFANHSIMRLA
SSANAKTIHIIDFGISDGFQWPSLIHRLAWRRGSSCKLRITGIELPQRGFRPAEGVIETG
RRLAKYCQKFNIPFEYNALAQWESIKLEDLKLKEGEFVAVNSLFRFRNLDETVAHVSP
RDTVLKLRIRKIPDVFIPIGLSGSYNAPFVTRFREVLFHYSSLFDMCDTNLTREDPMRV
MFEKFPYGREIMNVVACEGTERVERPESYKQWQARAMRAGFRQIPLEKELVQKLKLMVES
GYKPKFEDVDQDCHWLLQGWKGRIVYGSSIWVPPFFVYVGRATRVLIMDPNFSLSNGFEY
FDGNPNLLTDPMEDQYPPPSDTLLKYSBILMEESNGDYKQSMFYDSLALRKTEMLQQV
ITDSQNQSFSPADSLITNSWDASGSIDESAYSADPQVNEIMVKSMSFSDAESALQFKKGV
EEASKFLPNSDQWVINLDIERSERRDSVKEEMGLDQLRVKKNHERDFEEVRSSKQFASNV
EDSKVTDMFDKVLILLDGECDPQTLLDSEIQAIRSSKNIGEGKKKKKKKSQVDFRTLLT

HCAQAISTGDKTTALEFLQIRQSSPLGDAGQRLAHCANALEARLQGSTGPMIQTYYN
ALTSSLKDTAADTIRAYRVYLSSSPFVTLMYFFSIWMILDVAKDAPVLHIVDFGILYGFQ
WPMFIQSISDRKDVPRKLRTIGIELPQCGFRPAERIEETGRRLAEYCKRFNVPFEYKAIA
SQNWETIRIEDLDIRPNEVLAVNAGLRRLKLNQDETGSSENCPRDAVLKLRNMNPDVFIH
AIVNGSFNAPFFISRFKEAVYHYSALFDMFDSTLPRDNKERIRFEREFYGREAMNVIACE
EADRVERPETYRQWQVRMVRAGFKQKTIKPELVELFRGKLKKWRYHKDFVVDENSKWLLQ
GWKGRITLYASSCWVPA*

>G993 (6..1091)

CAAATATGGAATACAGCTGTGTAGACGACAGTAGTACAACGTCAGAATCTCTCTCCATCT
CTACTACTCCAAAGCCGACACGACGACGAGAGAAAGAACTCTCTTCTCCGCCGGCGACGT
CGATGCGTCTCTACAGAATGGGAAGCGGCGGAAGCAGCGTCGTTTGGATTACAGAGAACG
GCGTCGAGACCGAGTCACGTAAGCTTCCCTCGTCGAAATATAAAGCGTTGTGCTCAGC
CTAACGGAAGATGGGGAGCTCAGATTTACGAGAAGCATCAGCGAGTTTGGCTCGGTACTT
TCAACGAGGAAGAAGAAGCTGCGTCTCTTACGACATCGCCGTGAGGAGATTCCGCGGCC
GCGACGCCGTCACTAATCTCAAATCTCAAGTTGATGGAACGACGCCGAATCGGCTTTTC
TTGACGCTCATTCTAAAGCTGAGATCGTGGATATGTTGAGGAAACACACTTACGCCGATG
AGTTTGAGCAGAGTAGACGGAAGTTTGTTAACGGCGACGGAACGCTCTGGGTGGAGA
CGGCGACGTACGGAACGACGCTGTTTGTAGAGCGCGTGAGGTTTGTTCGAGAAGACTG
TTACGCCGAGCGACGTGGGAAGCTGAACCGTTTAGTGATACCGAAACAACACGCGGAGA
AGCATTTTCCGTTACCGCGATGACGACGCGGATGGGGATGAATCCGTCTCCGACGAAAG
GCGTTTTGATTAACTTGGGAAGATAGAACAGGGAAGTGTGGCGGTTCCGTTACAGTTACT
GGAACAGCAGTCAAAGTTACGTGTTGACCAAGGGCTGGAGCCGTTTCGTTAAAGAGAAGA
ATCTTCGAGCCGGTGATGTGGTTTGTTCGAGAGATCAACCGGACGACCGGCAATTGT
ATATCCACTGGAAGTCCGGTCTAGTCCGGTTCAGACTGTGGTTAGGCTATTCCGAGTCA
ACATTTTCAATGTGAGTAACGAGAAACCAAACGACGTCGAGTAGAGTGTGTTGGCAAGA
AGAGATCTCGGAAGATGATTTGTTTTCGTTAGGGTGTTCGAAGAAGCAGCGGATTATCA
ACATCTTGACAAATCTTTTTTTTGGTTTTTTTCTTCAATTTGTTTCTCTTTTCA
ATATTTTGTATTGAAATGACAAGTTGTAAATTAGGACAAGACAAGAAAAAATGACAACTA
GACAAAATAGTTTTTTGTTAAAAA

>G993 Amino Acid Sequence (domain in AA coordinates: 69-134)
MEYSCVDDSTTSSELSISTTPKPTTTEKKLSSPPATSMRLYRMGSGGSSVVLDSNGV
ETESRKLPSKSKYKGVVPQNGRWGAQIYEKHQVRVWLGTFNEEEAASSYDIAVRRFRGRD
AVTNFKSQVDGNDAESAFDLAHSKAEIVDMLRKHTYADEFEQSRKRFVNGDGKRSGLETA
TYGNDAVLRAREVLFKFTVPSDVGLNRLVIPKQHAEKHFPLPAMTTAMGMNPSPTKGV
LINLEDRTGKVVFRFRYSYWNSSQSYVLTKGWSRFVKEKNLRAGDVVCFERSTGPDRQLYI
HWKVRSSPVQTVVRLFGVNI FNVSNKPNDAVECVGKKRSREDDLFSLGCSKKQAIINI
L*

>G681 (1..804)

ATGGGGAGGACGACATGGTTCGACGTCGACGGGATGAAGAAAGGAGAGTGACGGCAGAG
GAAGACCAGAAGCTCGGCGTTACATCAACGAGCATGGCGTTGTGATTGGCGTTCCCTC
CCCCAAAGAGCTGGTTTGCAGAGATGTGGAAAGAGCTGCAGATTAAGGTGGCTTAATCT
CTAAAGCCTGGGATTAGAAGAGGCAAATTCCTCCTCAAGAAGAAGAAGAAATCATCCAA
CTTCATGCTGTTCTCGGAAACAGGTGGGCAGCCATGGCGAAGAAGATGCAGAATCGAACA
GACAATGATATCAAGAACCATTGGAACCTTGTCTCAAGAAAAGACTTTCGAGAAAGGGA
ATCGACCTATGACCCACGAGCCCATCATCAAACACCTCACCGTCAATACCACTAACGCA
GATTGTGGTAACCTTCCACCACGACGTCCCCGTCGACGACGGAAGCTCTCCTTCTCTC
GGCTCGTCTCGTCTTCTTAACAACTCGCCGAGGTATCTCATCTAGACAACATAGTCTC
GATAGGATCAAGTACATCTTGTGCAATTCAATAATCGAAAGCAGTGATCAAGCAAAAGAG
GAAGAAGAAAAAGAAGAAGAAGAAGAAGAGATTCAATGATGGGTGAGAAGATTGAC
GGTAGTGAAGGAGAAGATATTGAGATTTGGGGCAGGAGGAAGTTAGGCGTTTAATGGAG
ATTGATGCAATGGATATGTACGAGATGACTTCGTACGACGCTGTATGTACGAGAGTAGT
CACATACTTGATCATCTCTTTTACTTAATATAGTGTGACTGTGTGAGTGCATGCATGTT
>G681 Amino Acid Sequence (domain in AA coordinates: 14-120)

MGRTTWFVDVGMKKGEWTAEDQKLGAAYINEHGVCDWRSPLPKRAGLQRCGKSCRLRWLNY
LKPGIRRGKFTPQEEEEIIQLHAVLGNRWAAAMAKKMQNRDNDIKNHNWNSCLKKRLSRKG
IDPMTHEPIIKHLTVNTNADCGNSSTTTSPSTTESSPSSGSSRLNKLAAGISSRQHS
DRIKYLNSNIISSDQAKEEEEEKEEEEEERDSMMGQKIDGSEGEDIQIWGEEVRLME

IDAMDMYEMTSYDAVMYESSHILDHLF*

>G1482 (1..996)

ATGAAGATCAGGTGCGACGTCTGCGATAAAGAAGAAGCGTCGGTGTGTTTGCACGGCCGAC
GAAGCATCTCTCTGCGGCGGCTGCGACCACCAAGTCCACCACGCTAACAAACTCGCCTCT
AAACATCTCCGTTTCTCTCTCTTTATCCTTCTTCTTCCAACACCTCTCTCTCTCTGC
GACATCTGTGAGGATAAAAAAGCTCTGTTGTTCTGTCAACAAGATAGAGCTATTTTATGC
AAAGATTGCGATTTCATCGATCCACGCTGCGAACGAACACACAAAGAAACACGATAGGTTT
CTTCTTACAGGGGTTAAGCTCTCTGCAACATCGTCTGTTTACAAACCTACTTCGAAATCT
TCTTCTTCTTCTTCAAGCAACCAAGATTTCTCTGTCCCTGGATCATCAATCTCTAATCCT
CCTCCTCTCAAGAAACCTCTCTCAGCTCCTCCTCAGAGCAACAAGATCCAACCCCTTTCG
AAGATCAACGGCGGTGATGCGTCGGTGAATCAGTGGGGATCCACAAGCAGATTTCTGAG
TATTTGATGGATACGTTACCTGGTTGGCACGTTGAGGATTTCTCTGATTCCTCTCTTCTCT
ACTTATGGTTTCTCTAAGAGTGGTGTATGATGATGGAGTGTACCATATATGGAACCAGAA
GATGACAACAACACTAAGAGAAACAACAACAACAACAACAACAACAATACAGTG
TCACTTCCATCTAAGAATTTAGGGATTTGGGTCCCTCAGATTCCACAAACTCTTCTTCT
TCATACCCAAATCAATACTTTTCTCAAGACAACAACATACAGTTTGGGATGTACAACAAA
GAAACATCACCAGAAGTAGTGTCTTTTGTCCAATACAAACATGAAACAACAAGGACAG
AACAACAAGAGATGGTATGATGATGGTGGCTTCACTGTCCACAGATCACTCCTCCTCCT
CTTTCTCTAATAAAAAGTTTAGATCTTCTGGTAA

>G1482 Amino Acid Sequence (domain in aa coordinates: 5-63)
MKIRCDVCDKEEASVFCCTADEASLCGGCDHQVHHANKLASKHLRFSLLYPSSSNTSSPLC
DICQDKKALLFCQODRAILCKDCDSSIHAANEHTKKHDFLLTGVLKLSATSSSVYKPTSKS
SSSSSSNQDFSVPGSSISNPPPLKKPLSAPPQSNKIQPFKINGGDASVNQWGSTSTISE
YLMDTLPGWHVEDFLDSSLPTYGFSKSGDDDGVLPMPEPEDDNTKRNNNNNNNNNNTV
SLPSKNLGIWVPQIPQTLPPSSYPNQYFSQDNNIQFGMYNKETSPVVVSFAPIQNMKQGGQ
NNKRWYDDGGFTVPQITPPPLSSNKKFRSFW*

>G225 (157..441)

CTCTCTCTCTCACTCTTTCTTTTCCGAGAACCACAAAAAAGCTACTATTAATCC
TTCCCTCGTGAGGAAATCATTTCTTCTGTTTCTCGAGATTATTCTCTTCTCTCTCT
CTTTCTCTGTGTGTTTCGTGTCTTCAGATTAGTTCGATGTTTCGTTTCAGACAAGCGGAA
AAAATGATAAACGACGACGAGACAGAGCAAAGCCAAGGCTTCTTGTTCGGAAGAGGTG
AGTAGTATCGAATGGGAAGCTGTGAAGATGTGAGAAGAAGAAGATCTCATTTCTCGG
ATGTATAAACTCGTTGGCGACAGGTGGGAGTTGATCGCCGGAAGGATCCCGGACGAGCG
CCGGAGGAGATAGAGAGATATTGGCTTATGAAACACGGCGTCGTTTTTGGCAACAGACGA
AGAGACTTTTTTAGGAAATGATTTTTTTGTTTGGATTAAAGAAAATTTCTCTCTCTT
AATTCACAAGACAAGAAAAAAGGAAATGTACCTGTCTTGAATTACTATTTTGAATGT
ATAATATCTATATATATAAGAAGAAAAAATTGCTTAGGAATTT

>G225 Amino Acid Sequence (domain in AA coordinates: 39-76)
MFRSDKAEMDKRRRRQSKAKASCSEEVSSIWEAVKMSSEEBDLISRMVYKLVGDRWELI
AGRIPIGRTPPEIERIERYWLMKHGVVFANRRRDFFRK*

>G226 (10..348)

CCAGTAGTTATGGATAATACCAACCGTCTTCGTCTTCGTGCGGTCCAGTCTTAGGCAA
ACTAAGTTCACTCGATCCCGATATGACTCTGAAGAAGTGAGTAGCATCGAATGGGAGTTT
ATCAGTATGACCGAACAAGAAGAAGATCTCATCTCTCGAATGTACAGACTTGTGCGTAAT
AGGTGGGATTTAATAGCAGGAAGAGTCGTAGGAAGAAAGGCAAATGAGATTGAGAGATAC
TGGATTATGAGAACTCTGACTATTTTTCTCAAAACGACGACGTCTTAATAATTCTCCC
TTTTTTCTACTCTCTCTTAATCTCCAAGAAAATCTAAAATTGTAAAGAAATCAAAT
AAAAGCTTTCAATCATAAAAAGTAGAACAAATCTTGAATGTCTTCTCA

>G226 Amino Acid Sequence (domain in AA coordinates: 28-78)
MDNTNRLRLRRGPSLRQTKFTRSRDYSEEVSSIWEFISMTEQEEDLISRMVRLVGNRWD
LIAGRVVGRKANEIERIERYWIMRNSDYFSHKRRRLNNSPFFSTSPNLQENLKL*

>G9 (81..1139)

GTGTTTCTTCTTCTGCTAAAAGGTTATAATTTTTGTTTCTTGGTTTGGTGAGAATCTTC
AAGAACTGAAACAAAGAAATGGATTCTAGTTGCATAGACGAGATAAGTTCTCCACTT
CAGAATCTTTCTCGCCACCACCGCCAAGAAGCTCTCTCCTCCTCCGCGGCGGCTTAC
GCCTCTACCGGATGGGAAGCGGCGGAGCAGCGTCGTGTTGGATCCCGAGAACGCGCTAG
AGACGGAGTCACGAAAGCTACCATCTTCAAATACAAAGGTGTTGTTCTCAGCCTAACG

GAAGATGGGGAGCTCAGATCTACGAGAAGCACCAACGAGTATGGCTCGGGACTTTCAACG
AGCAAGAAGAAGCTGCTCGTTCTTACGACATCGCAGCTTGATAGATTCCGTGGCCGCGACG
CCGTCGTCAACTTCAAGAAGCTTCTGGAAGACGGCGATTTAGCTTTTCTTGAAGCTCACT
CAAAGGCCGAGATCGTCGACATGTTGAGAAAAACACTTACGCCGACGAGCTTGAACAGA
ACAATAAACCGGCAGTTGTTTCTCTCCGTGACGCTAACGGAAAACGTAACGGATCGAGTA
CTACTCAAACGACAAAGTTTAAAGACGTGTGAAGTTCTTTTCGAGAAGGCTGTTACAC
CTAGCGACGTTGGGAAGCTAAACCGTCTCGTGATACCTAAACAACACGCCGAGAAACACT
TTCCGTTACCGTCACCGTCACCGGCAGTGACTAAAGGAGTTTGTATCAACTTCGAAGACG
TTAACGGTAAAGTGTGGAGGTTCCGTTACTCATACTGGAACAGTAGTCAAAGTTACGTGT
TGACCAAGGGATGGAGTCGATTTCGTCAAGGAGAAGAATCTTCGAGCCGGTGATGTTGTTA
CTTTCGAGAGATCGACCGGACTAGAGCGGCAGTTATATATTGATTGGAAGTTTCGGTCTG
GTCCGAGAGAAAAACCGGTTCAAGGTGGTGGTTCGGCTTTTCGGAGTTGATATCTTTAATG
TGACCACCGTGAAGCCAAACGACGTCGTGGCCGTTTTCGGTGGAAAGAGATCTCGAGATG
TTGATGATATGTTTGCCTTACGGTGTTCGAAGAAGCAGGCGATAATCAATGCTTTGTGAC
ATATTTCTTTTCCGATTTTATGCTTTTCGTTTAAATTTTTTTTTTGTCAAGTTGTGT
AGGTTGTGATTATGCTAGGTTGTATTTAGGAAAAGAGATAAGACC

>G9 Amino Acid Sequence (domain in AA coordinates: 62-127)

MDSSCIDEISSSTSEFSATTAKKLSPPPAALRLYRMGSGSSVLDPENLETESRKL
PSSKYKGVVPQNGRWGAQIYEKHQVWLGTFFNEQEEAARSYDIAACRFRGRDAVVNFKN
VLEDGDLAFLEAHSKAEIVMLRKHTYADELEQNNKRLFLSVDANGKRNGSSTQNDKV
LKTCEVLFEKAVTPSDVGLNRLVIPKQHAKEHFPLPSPSPAVTKGVLINFEDVNGKVWR
FRYSYWNSSQSYVLTGWSRFVKEKNLRAGDVVTFERSTGLERQLYIDWKVRSQPRENPV
QVVVRLFGVDIFNVTTVPKNDVVAVCGGKRSRDVDDMFALRCSKKQAIINAL*

>G1040 (51..863)

CTTTGATCTCCACTATTTAAGTAGACAAGAATCATAAAGAAAATAGTGAGATGATGATGT
TAGAGTCAAGAAACAGTATGAGAGCTTCAAACCTCAGTCCCAGATCTGTCTCTTCAGATCA
GTCTTCTTAATATCACGCCGAAAACCTCTTCACGGCGGTGACCGGAGCTCCACAAGCA
GTGATTCTGGAAGCAGCCTCAGTGACCTGAGCCATGAGAACAACCTTCTTCAACAAACCTC
TCTTGAGCTTAGGATTTGACCATCATCATCAAAGGCGCTCAAACATGTTCCAACCTCAA
TCTACGGTCGAGATTTCAAGAGAAGCTCATCATCAATGGTTGGTCTTAAACGAAGCATT
GTGCTCCAAGAATGAGATGGACTTCTACTCTTCATGCTCACTTCGTCCATGCTGTTCAAC
TTCTTGCGCGCCATGAAAGAGCAACGCCTAAATCAGTGTGGAGCTCATGAATGTGAAGG
ATCTAACCTTAGCTCATGTCAAGAGTCACTTGCGAGATGTATAGAACAGTGAAATGCACTG
ATAAAGGATCACCAGGAGAAGGAAAGGTAGAGAAAGAGGCAGAGCAGAGGATAGAGGACA
ATAATAAATGAAGAAGCTGATGAAGGAAGTACACAAATTCGCCAAACTCATCATCTG
TGCAAAAGACCCAAAGAGCTTCAATGGTCATCGACAAAGGAAGTATCTAGGAGCATATCTA
CACAAGCATATTTCACTTGGGAACAACCTCATCACACTAAGGCCAATGAAGAGAAAGAGG
ATACCAACATTCATCTCAATTTGGATTTACATTGGGCGGCCTAGTTGGGGGATGGAATA
TGCGGAACCCCTCCAGTGATTTAAACCCTTCTCAAGTGCTAATTGCCTTAAGCTACAACAAA
TAAGTCAGCTTAGGTTACCAGTTTAAACATAATTTAACTTGTTTTGATCATATGAGCTT
CGGAAGAATCATATTATCATCATATATGAACCTTCTTCCAAGAATGTTCTATGAGTTTTT
TGATATGTATAATCAAGAGAATCGTTTGAAGTAAAAA

>G1040 Amino Acid Sequence (domain in AA coordinates: 109-158)

MMLESRNSMRASNSVPDLSLQISLPNYHAGKPLHGGDRSSTSSDSGSSLSDLSHENNFF
NKPLLSLGFDDHHHQRSSNMFPQPIYGRDFKRSSSMVGLKRSIRAPRMRWTSTLHAHFVH
AVQLLGGHERATPKSVLELMNVKDLTLAHVKSHLQMYRTVKCTDKGSPGEGKVEKEAEQR
IEDNNNNEEADEGTDNTPNSSSVQKTQRASWSSTKEVSRISSTQAYSHLGTTHHTKANE
EKEDTNIHLNLDFTLGGGLVGGWNMRNPPVI*

>G2114 (64..1311)

ATAAAACGAAACCTATACATATAAACTAAGAGCGAGAAAGACAGCTAGAGAGAGAGAGA
GAGATGAAGAAATGGTTGGGATTTTCATTGACACCTCCTTTGAGAATCTGCAATAGTGAA
GAAGAAGAACTTAGGCATGACGGTTCCGATGTTTGGAGATATGATATTAACCTTTGATCAT
CATCATCATGATGAAGACGTTCCAAAGGTGGAAGATCTCCTCTCAAACCTCTCATCAAACC
GAGTATCCTATAAACCATAACCAAAACCAATGTCAACTGCACCACTGTGGTTAACAGGTTA
AACCACCCGGTTACCTTCTCCACGACCAACCGTAGTTACACCACATTACCCGAACCTA
GATCCGAACCTTAGCAATGATTATGGAGGTTTTGAGAGGGTCGGTTCGGTCTCGGTTTTC
AAATCTTGTTAGAGCAAGGCATCCAGCATTCCCACTCTCGAGTCATTACGTTACTGAA

GAGGCTGGTACGAGCAATAATATTAGTCATTTTAGTAACGAAGAGACTGGTTATAACACC
AATGGCTCAATGCTATCATTGGCTTTGAGCCATGGGGCTTGTCTGATTTGATCAACGAA
TCGAATGTATCCGCACGGGTGGAAGAACCAGTTAAGGTAGATGAGAAGCGGAAGAGATTG
GTTGTTAAACCTCAGGTAAGGAATCCGTTCCCTCGGAAGTCGGTTGATAGTTATGGACAA
AGAATTCTCAGTATCGTGGAGTTACAAGGCATAGATGGACAGGGAGATATGAAGCTCAC
TTATGGGATAATAGCTGTAAGAAGGAGGGACAGACAAGGAGAGGAAGACAAGTGATCTT
GGAGGGTATGATGAGGAGGAGAAAGCAGCGAGGGCATATGATTTAGCGGGCTCTGAAGTAT
TGGGGTCCCTACCACTCACTTAAATTTCCCTTTGAGTAATTACGAAAAGGAGATCGAGGAA
CTCAATAACATGAATCGGCAAGAATTTGTTGCCATGTTGAGGAGGAATAGCAGCGGGTTT
TCGAGGGGAGCTTCCGTGTATAGAGGAGTTACAAGGCATCATCAACATGGAAGGTGGCAA
GCCAGAAATTGGAAGAGTTGCTGGAACAAGGACTTGTACCTTGAACATTTAGCACGCAA
GAAGAAGCAGCGGAGGCGTACGATATCGCGGCAATTAAATTCAGAGGCCTAAACGCTGTA
ACCAATTTTCGATATAAATAGATATGACGTGAAGAGGATATGTTCAAGCTCAACGATTGTT
GATAGCGACCGAGGCCAAACATTTCCACAGCTCTGGCGCCGCCACTAACCGACACCG
TAACTCCTCGCCGGAGAGACTATTTCCACGTACGGTTGGTTTGGAGGAAATAAGTTCGTC
CAGTCTGTTTAAATCATTTATGGTTAATAAACAATATATTCCTAAGTAATTGAGGCCGGTC
TACATATATACAACTTTTTTAGCAAATTAAGTTATCAGAATCCACTATATATTATTCTCT
>G2114 Amino Acid Sequence (conserved domain in AA coordinates: 221-297, 323-393)
MKKWLGLFSLTPPLRICNSEEELRHDSGVWRYDINFDDHHHDEDPKVEDLLSNHQTE
YPINHNQTNVNCCTTVNRLNPPGYLLHDQTVVTPHYPNLDPNLSNDYGGFERVGSVSVFK
SWLEQGTAPFLSSHYVTEEAGTSNNISHFSNEETGYNTNGSMLSLLSLSHGACSDLINES
NVSARVEEPVKVDEKRRKRLVVKPQVKESVPRKSVDSYQRTSQYRGVTRHRWTGRYEHL
WDNSCKKEGQTRRRGRQVYLLGGYDEEKAARAYDLAALKYWGPTTHLNFPLSNYEKEIEEL
NNMNRQEFVAMLRNSSGFSRGASVYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFTSTQE
EAAEAYDIAAIKFRGLNAVTFNFDINRYDVKRICSSSTIVSDQAKHSPTSSGAGH*
>G450 (65..751)
GAGTTATCGAGAGAGAGAGAAAAACATATTCTGATTTAAGACATATATAGACAGCAAGAAG
AGATATGAACCTTAAGGAGACGGAGCTTTGTCTTGGCCTCCCCGGAGGCCTGAAACCGT
TGAAAGTCCGGCCAAGTCGGGTGTTGGGAACAAGAGAGGCTTCTCCGAGACCGTTGATCT
CAAACCTTAATCTTCAATCTAACAAACAAGGACATGTGGATCTCAACACTAATGGAGCTCC
CAAGGAGAAGACCTTCCCTTAAAGACCCTTCTAAGCCTCCTGCTAAAGCACAAAGTGGTGGG
TTGGCCACCGGTGAGGAACCTACCGGAAAAATGTTATGGCTAATCAGAAGAGCGGCGAAGC
AGAGGAGGCAATGAGTAGTGGTGGAGGAACCGTCGCTTTGTGAAGGTTTCCATGGATGG
AGCTCCTTATCTTCGGAAGGTTGACCTCAAGATGTACACCAGCTACAAGGATCTCTCTGA
TGCCTTGGCCAAAATGTTCAAGCTCCTTTACCATGGGGAGTTATGGAGCACAAAGGGATGAT
AGATTTTCATGAACGAGAGTAAAGTGATGGATCTGTTGAACAGTTCTGAGTATGTTCCAAG
CTACGAGGACAAAGATGGTGAAGTGGATGCTCGTTGGTGTATGTCCTCGGCGATGTTTGT
CGAGTCATGCAAAAGCTTTGCGCATATGAAAGGATCCGAAGCAATTGGACTTGCTCCAAG
AGCAATGGAGAAGTTCAAGAACAGATCATGAACAAAAAAGAGGACAAATATGCATTG
ATTTTTTTTTTTTTTGGTATTGTTATGATCATGTGTTTTAATTAAATATAGGAAGGATA
TAGGAAAATATAATTGTTTACAAAAAATAACTTTAAATATAGTCTTTTTTTTTTTTGA
AATTAGTCTGTGTTTTGTTTTTCATCTCTTAATTAGTAGAAATCATTTTTTTAATATGTAA
TTGTGATAGTAAATCTATAGAGTTCGTA
>G450 Amino Acid Sequence (domain in AA coordinates: TBD)
MNLKETELCLGLPGGTETVESPAKSGVGNKRGFSETVDLKLNLQSNKQGHVDLNTNGAPK
EKTFLKDPKPPAKAQVVGWPPVRNYRKNVMANQKSGEABEAMSSGGTVAFVKVSMGDA
PYLRKVDLKMYSYKDLSDALAKMFSSFTMGSYGAQGMIDFMNESKVMDDLNSSEYVPSY
EDKDGDMVLVGDVPMFVESCRLRIMKGSEAIGLAPRAMEKFKNRS*
>G584 (40..1809)
AAAAAGTCTTCTCTTTTATAACTACGTCAGAGAACTGTTATGTCTCCGACGAATGTTCAA
GTAACCGATTACCATCTCAACCAATCAAAAACGGATACAACAAATCTCTGGTCAACCGAC
GACGATGCATCGGTAATGGAAGCTTTTCATCGGCGGCGGCTCCGATCAITCTTCTCTTTTT
CCTCCACTTCTCCTCCTCCTTCTCCTCAAGTCAACGAAGATAATCTCCAGCAACGCTCTC
CAAGCTTTAATCGAAGGAGCAACAGAGAACTGGACTTACGCCGTGTTCTGGCAATCATCT
CACGGTTTCGCCGGAGAAGACAACAACAACAACACAGTGTGTTAGGTTGGGGAGAT
GGTTATTACAAAGGAGAAGAAGAGAAGTCTAGAAAGAAGAAATCCAGCTAGTGCA
GCTGAACAAGAGCATCGTAAGAGAGTGATTAGAGAGCTCAACTCTTAATCTCCGGTGGT

GTAGGAGGAGGAGATGAAGCTGGAGATGAAGAAGTTACAGATACTGAATGGTTCTTCTTA
GTTTCAATGACACAGAGCTTTGTCAAGGGTACTGGTTTACCTGGTCAAGCTTTCTCAAAT
TCAGACACGATTGGTTATCTGGTTCTAATGCTTTAGCTGGATCAAGTTGTGAGAGAGCT
CGTCAAGGTCAGATTTATGGGTTACAAACAATGGTGTGTGTAGCGACAGAGAATGGTGTG
GTTGAGCTTGGTTTCGTCGGGAGATTATTCATCAAAGTTCAGATCTTGTGATAAAGTTGAC
ACCTTTTTCAATTTTAAACAATGGTGGTGGTGAATTTGGTTCTTGGGCGTTAATTTGAAT
CCAGATCAAGGAGAGAATGATCCAGGTTTGTGGATTAGTGAACCTAATGGTGTGACTCT
GGTCTTGTAGCTGCTCCGGTGATGAATAATGGTGGAAATGACTCAACTTCTAATTCTGAT
TCTCAACCAATTTCTAAGCTTTGTAATGGAAGCTCTGTTGAAAACCTAACCTTAAAGTT
CTGAAATCTTGTGAAATGGTGAATTTCAAGAATGGGATTGAGAATGGTCAAGAAGAAGAT
AGTAGTAATAAGAAGAGATCACC GGTTTTCGAATAATGAAGAAGGGATGCTTTCTTTTACC
TCTGTTCTTCCATGTGACTCGAATCACTCTGATCTTGAAGCTTCAGTGGCTAAAGAAGCT
GAGAGTAACAGAGTTGTGGTTGAACCGGAGAAGAAACCGAGGAAACGAGGGAGAAAACCG
GCGAATGGAAGAGAGAGAGCCTTTGAATCATGTAGAGGCAGAGAGACAGAGAAGAGAGAAG
TTGAATCAGAGATTCTATTCTTTAAGAGCTGTGGTTCCTAATGTGTCTAAGATGGATAAA
GCTTCTCTATTAGGAGATGCTATTTTCGTATATCAGTGAGCTTAAGTCTAAGTTGCAAAAG
GCTGAATCTGATAAAGAGAGTTGCAGAAGCAGATTGATGTGATGAATAAAGAAGCGGGA
AATGCGAAAAGTTTCGGTAAAAGATCGAAAATGTTGAATCAAGAATCAGAGTGTGTGATA
GAGATGGAGGTTGATGTGAAGATTATTGGTTGGGATGCAATGATAAGGATTCAATGTAGT
AAGAGGAATCATCTGGTGCTAAGTTCATGGAAGCACTTAAGGAGTTGGATTGGGAAGTG
AATCATGCGAGTTTATCGGTAGTGAATGATCTTATGATCCAACAAGCGACTGTGAAAATG
GGGAATCAGTTTTTTCACGCAAGATCAACTCAAGGTTGCTCTAACGGAGAAAGTTGGAGAA
TGTCATGAATTGAAGTCAGCATCTTTAGGGCTAATACACCGGAGAATACTGCGAAAAGT
CGAAAACAACGATCATAGTATAAGCCGCGGTAAAAAGTGTAAACCTTTTACACAAGTTT
CTCTAGTGAATGTAGTTGTAACTCTATTGTGTAAAGGTAATTTGTAGTACCCACTTGT
TGCTATTGAATGCTTGTGTAGAGAGGATTCTTAGTGTAGTATATGATTAGGTTGGGGTTG
TTGTTTCATGAGATAAATAAATGTGTTGATCAATGGTTAAGTCTTTGGTTTGTGGTGT
ATGTATGTAATAAGGCTTTTGTGTAGAAATAAGACAAATGGGACTGAAGTTGGAGTTTAA
AA

>G584 Amino Acid Sequence (domain in AA coordinates: 401-494)
MSPTNVQVTDYHLNQSKDTTTLNWSDDASVMEAFIGGGSDHSLFPPLPPPPLPQVNE
DNLQORLQALIEGANENWTYAVFWQSSHGFAGEDNNNNNTVLLGWGDGYKKEEEKSRKK
KSNPASAAEQEHRKRVIRELNSLISGGVGGGDEAGDEEVDTEWFFLVSMQTQSFVKGTGL
PGQAFNSDFTIWLSGSNALAGSSCERARQGQIYGLQTMVCVATENGVELGSSEI IHQSS
DLVDKVDTFNFNNGGEGFWSWAFNLNPDQGENDPGLWISEPNGVDSGLVAAPVMNNGGN
DSTNSDSQPISKLCNGSSVENPNPKVLKSCMVNFKNIGENGQEDSSNKKRSPVSNNE
EGMLSFTSVLPDCSNHSDLEASVAKEAESNRVVVEPEKKPRKRGRKPANGREEPLNHVEA
ERQRREKLNQRFYSLRAVVPNVSKMDKASLLGDAISYISELKS LKQKABSDKEELQKQID
VMNKEAGNAKSSVKDRKCLNQESSVLIEMEVDVKIIGWDAMIRIQCSKRNPAGAKFMEAL
KELDLEVNHASLSVNDLMIQQATVKMGNQFFTQDQLKVALTEKVGECP*

>G668 (1..1056)

ATGGGAAGACCACCTTGCTGTGAAAAGATTGGAGTGAAGAAAGGGCCATGGACACCAGAG
GAAGACATCATCTTGGTTTCTTACATCCAAGAACATGGTCCTGGAACTGGAGATCTGTC
CCAACACACACAGGTTTAAAGATGTAGCAAGAGCTGCAGATTGAGATGGACTAATTATCTT
CGACCCGGTATTAAGCGTGGAAATTTTACTGAGCATGAAGAGAAGACAATTGTTTCATCTT
CAAGCCCTTTTAGGCAACAGATGGGCAGCCATAGCATCATACCTTCCAGAAAGGACAGAC
AATGATATAAAGAATAATTGGAACACTCACTTGAAGAAGAAGCTCAAAAAGATTAAATGAA
TCTGGTGAAGAAGATAATGATGGTGTCTCTTTCATCAAACACTAGTTCAAAAAGAACCAT
CAAAGCACTAACAAAGGTCAATGGGAAAGAAGACTTCAGACAGACATTAACATGGCAAAA
CAAGCTCTTTGTGAGGCCTTGTCTTTAGACAAAACATCATCCACTCTTTCATCATCTTCA
TCATTACCGACACCAGTAATCACACAACAAAACATCCGTAACTTCTCATCAGCTTTGCTT
GACCGTTGTTATGATCCATCTCTTCTTTCATCTACCACAACCACCACTACAAGCAAC
ACTACTAATCCATACCATACCGATCAGGGGTATATGCGTCAAGTGCTGAGAACATCGCCCCGTTG
CTTCAAGATTTTCATGAAAGACACACCAAGGCTTTAACTTTATCATCTTTCATCTCCGGTT
TCAGAGACTGGACCACTCACTGCTGCAGTCTCGGAAGAAGGTGGAGAAGGGTTTGAACAA
TCTTCTTTCAGCTTCAATTCAATGGACGAAACTCAAACTTGACTCAGGAGACAAGCTTC
TTCCATGATCAAGTGATCAACCCGGAAATAACAATGGACCAAGATCATGGTCTAATATCA

CAAGGGTCTCTGTCTTTGTTTGTGAGAAATGGTTATTGTATGAGCAAAGCCACGAGATGGTT
GGTATGGCACTAGCAGGACAAGAAGGGATGTTCTAG
>G668 Amino Acid Sequence (domain in AA coordinates: 13-113)
MGRPPCCEKIGVKKGPWTPEEDIILVSYIQEHGPGNWRVPTHTGLRCSKSCRLRWNYL
RPGIKRGNFTEHEEKTIVHLQALLGNRWAAIASYLPERTDNDIKNYWNTHLKKLKKINE
SGEEDNDGVSSSNTSSQKNHQSTNKGQWERRLQTDINMAKQALCEALSLDKPSSTLSSSS
SLPTPVITQQNIRNFSSALLDRCYDPSSSSSTTTTTSNTNPNYPGVSASSAENIARL
LQDFMKDTPKALTSSSSPVSETGPLTAAVSEEGGEGFEQSFSSFNMDDETQNLQETSF
FHDQVIKPEITMDQDHLISQGSLSLFEKWLFDQSHMVMALAGQEGMF*
>G1050 (23..1582)
TTCCCCATTTCAGAAAATCAAAATGGGTGGTGGTGGTGATACAACAGATACCAATATGAT
GCAGAGAGTTAATCTTCTTCTGGTACATCGTCTTCTTCGATCCCTAAACACAATCTTCA
CTTGAATCCTGCTCTTATCCGCTCTCACCATCACTTCCGTCACCCCTTTCACCGGAGCTCC
TCCACCGCCGATTCCACCCATTCTCCTTACTCTCAGATCCCGCGACTTTACAACCTAG
ACATTCTCGCTCTATGTGCGAACCGTCTTCTTCTTCTCCTTTGATTCAATGCGCGCGTT
AAATCCTTCTGCTCCGTCGGTTCGGTGTGGTGGAGGAGAAAACCGTGCCGCGATTAG
TCCTTCGTTGCTCCGTCACCGTTACGATGTGTCAATCTTCTAGCTCTAGGAACGCCGG
AGATGGAGAGAATCTACCTCCGAGAAAGTCGCATAGGCGTTCGAATAGTGATGTTACTTT
TGGGTTTAGTTCAATGATGTCTCAGAATCAAAAGTCTCCTCTTCTGAGTTCTTTGGAGAG
ATCGATCTCTGGTGAAGATACATCAGATTGGTCTAATTTGGTGAAGAAAGAACCGAGAGA
AGGCTTCTACAAGGAAGAAAACAGAGGTTGAAGCAGCTATGGACGATGTTTTACGGC
TTATATGAATCTTGATAACATTGATGTCTTGAATTCTTTGGAGGTGAAGATGGCAAGAA
TGGAATGAGAATGTGGAGGAGATGGAGAGTAGTAGAGGTAGTGGTACAAAGAAGACGAA
TGGTGAAGTAGTAGTGATTCTGAAGGAGATAGCAGTGCAGTGGGAATGTGAAGGTTGC
GTTGAGTTCTTCTTCTTTCAGGCGTGAAGAGAAGAGCAGGTGGAGATATTGCTCTACTGG
TAGACATTACAGGAGTGTTCCTATGGACAGTTGTTTCATGGGGAAGTTGAATTTCCGCCGA
CGAATCATCGCTAAAGCTTCCGCTTCTTCTCATCAGCTAAAGTTTCCCAACCAATTTCAGG
TGAAGGAATTCAAGTGCTTATAGTGTGAATTTGGAACAGTGAGTTTACTGCAGCTGA
AATGAAGAAGATTGCAGCTGATGAGAACTCGCTGAGATTGTAATGGCTGACCTTAAGCG
TGTTAAAGAATCTTGGCGAACCGCGTATCTGCTGCACGTTCAAAGGAGCGGAAGACGCG
ATACATGGCAGAGTTGGAACACAAGGTGCAGACACTTCAGACTGAAGCTACTACATTATC
GGCTCAGCTCACACATTTCAGAGAGATTCTATGGGGTTGACAAACAGACAGTGAGCT
GAAGTTTCGCTTTCAGCTTATGGAGCAGCAAGCACAACCTCCGCGATGCTCTGTCTAGAGAA
ACTGAATGAAGAAGTCCAGCGGTTGAAACTGGTGTAGGGGAGCCGAACCGCAGGCAAAG
TGGGAGCAGCAGCAGCAATCAAAGATGTCACTAAACCCGAGATGTTTCAGCAGCTTAG
CATAAGTCAGTTACAACACCAACAGATGCAGCATTCCAATCAGTGTAGCACAATGAAAGC
AAAGCACACTTCAAACGACTAGGGTAAGTAAACTGCGATCCGAGTTGTCTAGTTACAT
ATATGATAAGAATCTTTTGTGCAGAGTTCTGTTTTTGAAGTTTTAAAGAAACATATATA
AAGATTATGTCGGGAAATTTGATCATATTTCTGAAACATACACATATATATAGTGG
TAATGGAGGACTTTCTTCTGGACCA
>G1050 Amino Acid Sequence (domain in AA coordinates: 372-425)
MGGGGDTTDTNMQRVNSSSGTSSSSIPKHNHLNLPALIRSHHFRHPFTGAPPPPIPIPI
SPYSQIPATLQPRHSRMSQPSFFSFDLPLNPSAPSVSVSVEEKTGAGFSPSLPPSP
FTMCHSSSRNAGDGENLPPRKSHRRSNSDVTFGFSSMMSQNKSPPLSSLERSISGEDT
SDWSNLVKKEPREGFYKGRKPEVEAMDDVFTAYMNLNDIDLNSFGGEDGKNGNENVEE
MESSRSGTGKKTNGSSSDSEGDSSASGNVVKVALSSSSSGVKRRAGGDIAPTGRHYRSVS
MDSFCFMGLNFGDESSLKLPPSSSAKVSPTNSGEGNSSAYSVEFGNSEFTAAEMKKIAD
EKLAEIVMADPKRVKRI LANRVSAARSKERKTRYMAELEHKVQTLQTEATTLAQLTHLQ
RDSMGLTNQNSLKFRLQAMEQQQLRDALSEKLNVEVQRLKLVIGEPNRRQSGSSSSSES
KMSLNPMPFQQLSISQLQHQQMQHSNQCSTMKAKHTSND*
>G1463 (199..1209)
TATCCTTCGCAAGACCTTCTCTATATAAGGAAGTTCATTTCAATTTGGAGAGGACACGC
TGACAAGCTGACTCTAGCAGATCTGGTACCGTCGACAGTTTGAGATTTGCTTCATCCGGT
TTTTTTATTTCTGCAAAATATGTCACTCTCTCCCATTTTGTTTCATATATAATATGTTG
AAGTTTGATCAACTTAGTATGCGTTTCTTTTCTCTCTAGTTCCTCTGTTTCTGGTCTGA
TTTAGTTTCGTTATGGCGGACACACTGCTCAACGCAGAAGACGAAGTAATAATCTCACGT
TATCTGAAGCCTATGATCGTTAACAGAGTATCATGGCTGATCTCTTCATCGAAGACGCA

GACGTGTTCAACAAGGATCCATATGTGAAGTTCATGCTGAGATCCCTAGCTTCGTGATC
GTTAAACCACGAACAAAGGCTTGTGGTAAAACCGATGGATGTGATTCGGGTTGCTGGAGG
ATCATTGGTCGTGATAAGCTGATAAAGTCGGAGGAGACTGGTAAGATTCTAGGGTTCAAG
AAGATACTCAAGTTCTGCCTAAAGTGGAAACCTAGAGAATACAAGAGAAGTTTGGTAATG
GAAGAGTATAGGCTTACCAATAACTTCAACTGGAAGCAAGATCATGTGATTTGCAAGATT
CGGCTTTTGTGTTGAAGCAGAAATTAGTTTCTTGCTAGCCAAGCATTCTACACTACATCA
GACTCACTTCCTCGAAATGTGCTGTTGCCAGCTTATGGATTCTGTTACCAGATAAAACAA
GAGGAGGACGAATTTTATCCGGTGACGATAATGATTTTCAAGGAAAAGATTGGCCTAGC
TACGTTACCAACAACGTGTATTGTCTGCATCCATCGAGCTTGTGAATGTTACGATGGG
AAGTTTCATGATAACGGAATCTGCATCTTCGCTAACAGGACTTGTGGTGTAACCGATAAA
TGCAATGAAGGTTACTGGAAGATTAAGCACCGTGAGAAGCTGATCATGTACGGTACGGG
CAGACCATTGGTTGGAAGAAAGTTTTTCAGTTTTATGAAACGGAGAAAAGAACATTTT
GGTAATGGAGAAGAAGTGAAGGTAACCTGGACTCTAAAAGAGTATAGGCTTACCAGAAAA
ATGAACAAGATAAAGTGGTGTGCGTTATCAAGTATAAGGTAAAGTGTTCACCGAGGATA
ACTAGCTAGGGAAGCTTACTCTTGGTTTCATGATCGATGCGACCGCTCTAGACAGGCCTC
GTACCGGATCCTCTAGCTAGAGCTTTTCGTTTCGTATCATCGGTTTCGACAACGTTTCGTC
>G1463 Amino Acid Sequence (conserved domain in AA coordinates:9-156)
MRFFSLVPLFLGRFSFVMDTLNNAEDEVIIISRYLKPMIVNRVSWPDLFIEDADVFNKD
PYVKFHAIEPSFVIVKPRTKACGKTDGCDSGCWRIIGRDKLIKSEETGKILGFKKILKFC
LKWKPREYKRSVLMBEYRLTNFNWQDQHVICKIRLLFEAEISFLLAKHFYTTSDSLPRN
VLLPAYGFCSDPKQEEDEFYPVTIMISEGKDWPSYVTNNVYCLHPSSELVNVHDGKFHDNG
ICIFANRTCGVTDKCNENYWKIKHREKLIMSRYGQTIGWKKVFQFYETEKERHFNGGEEV
KVTWTLKEYRLTRKMNKNKVVCVIKYKVKCLPRITS*
>G1944 (236..1306)
TCGACCTTCTAATTTCCAACCTCTGTTCTTAGCAATATATTTTTCTCCAAAAATAATT
CTCAGTTTGATTTTCTTCTCTAGCTCTTAAGTATATTTCTTTGTTGTTATTTATCTTTT
AATCCTTTAATCTCATCTTTGTTTATCTTTAATCAAAACCCAAAATTTACATGGGTTCTT
GAAAATCTAGAGAAATAAAGGAAACATAACAAAAATAGAAAGAAAAAGAAGCTAATGGT
CTTAAATATGGAGTCTACCGGAGAAGCTGTTAGATCAACCACCGGTACGACGGTGGTAT
TACGGTGGTTAGATCCGACGCGCGCTCAGATTTCCACGTAGCTCAAAGATCAGAAAGCTC
AAACCAATCTCCACCTCTGTCACTCCTCCTCCACCACAGCCATCGTCTCATCACACAGC
TCCTCCGCGCGTGCAAATTTGACGCGTGACGACTACGACTACGACGCGCGCGATGGAAGG
TATCTCCGGTGGACTGATGAAGAAGAAGCGTGACGCGCAAGGAAGTATGGACCGGACGG
GACTGTTGTAGCGTTATCTCTAAACCGATTTCATCAGCGCGCGCGCGTCCGATCTTCC
GCCGCGGAGTTACACGTCATCGATTTCTCCGCTTCTGAGAAACGTAGCAAAGTGAAACC
AACGAACCTCGTTAACAGAACAAAGTATCATCACCAAGTTGAGAATTTGGGTGAATGGGC
TCCTTGCTCCGTCGGTGGTAATTTACACCTCATATAATCACAGTCAACACCGGCGAGGA
TGTAACAATCGATAAATCTCGTTTTTCGCAACAAGGACCTCGCTCTATTGTGTCTGTCT
AGCAAACGGTGTATTTCAGCGTTACACTTCGTGAGCCAGATTCTCTGGCGGCACATT
GACATACGAAGTCCGTTTGAAGATATTATCATTATCCGGGTCATTATGCCTAATGATTC
AGGCGGAACACGAAGTAGAACGGGAGGAATGAGTGATCGTTAGCAAGTCCCGATGGACG
TGTAGTAGGCGGTGGCCTCGCCGGTTTACTAGTAGCCGCGAGTCCGGTTCAGGTGGTGT
AGGAAGTTTTTTAGCGGGCACTGACCATCAAGATCAGAAACCGAAAAAGAACAAACATGA
TTTCATGTTGTCGAGTCTTACCGTGCATTCCTATCTCTAGTGACGCTGATCACCGGAC
AATCCATTCCGTCTCGTCTCTTCCGGTCAATAATAATACATGGCAGACTTCTTTAGCTTC
CGATCCAAGAACAAAGCATAACCGATATTAATGTCAATGTAAGTGAATCCAATCTTTCT
CTGTATTTTCTGTTAACAAGTTTGATTTGGTTGTTTATCTACATTAGGATTTTACTAAAA
TGGTAGTATTATTATAGGTTTTAGGGTCTTTATTTTGGTCCACTGTTGTCACTTGTA
GGATA
>G1944 Amino Acid Sequence (domain in AA coordinates:87-100)
MVLNMESTGEAVRSTGNDGGITVVRSDAPSDFHVAQRSESSNQSPSTVPPPPQPSHH
TAPPLQISTVTTTTTTAAMEGISEGLMKKRGRRPRKYGPDTVVALSPKPISSAPAPSH
LPPPSHHVIDFSASEKRSVKPTNSFNRTKYHHQVENLGEWAPCSVGGNFTPHIITVNTG
EDVTMKIISFSQQGPRISCVLSANGVISSVTLRQPDSSGGTLTYEGRFEILSLSGSFMFN
DSGSTRSRTGMSVSLASPDGRVVGGGLAGLLVAASPVQVVVGSFLAGTDHQDQKPKKNK
HDFMLSSPTAATPISSAADHRTIHSVSSLPVNNNTWQTSLASDPRNKHTDINNVNT*
>G2383 (37..990)

GACCTCTTTGATCCCTTCATTCCCATCAAACAACCATGTTTCCTTCTTTTCACTACTCAC
ATTCAAAGCCCTAATTCTCACCATCACTACTCTTCGCCTTCTTTTCCCTTCTCTCCGAT
TTTCTTGAGAGTTTGTGATGAATCCTTCTTGATAAACCAATTCTTGTTACAGCAGCAAGAT
GTAGCAGCAAATGTTGTTGAATCTCCTTGGAATTTTGAAGAAGCTTGAGCTTAAGAAG
AAGAATGAGAAGTGTGTTGATGGAAGCACCTCACAAGAGGTTCAATGGAGAAGGACGGTC
AAAAAAGGGACAGGCATAGTAAGATCTGCACGGCTCAAGGTCCTAGAGACCGGAGGATG
AGGCTGTCTCTCAGATTGCTCGCAAGTTTTTCGATCTTCAAGACATGTTGGGTTTCGAC
AAGGCGAGCAAGACGATTGAATGGCTTTTCTCCAAATCAAAGACTTCCATCAAACAACCTT
AAAGAAAGAGTGGCTGCATCGGAAGGAGGAAAGGATGAACATCTCCAGGTTGATGAA
AAGGAAAGGATGAGACACTGAAGTTGAGAGTCTCAAAGAGAAGAACAAGACTATGGAG
AGCTCTTTTAAAGACTAAAGAGTCGAGAGAGAGAGCTAGAAAGCGAGCAAGAGAGAGAACA
ATGGCAAAGATGAAGATGAGATTATTTGAGACCTCGGAAACAATTTAGATCCTCATCAA
GAACTAGAGAGATCAAGATAACCAATGGTGTAACAATTACTAGAAAGGAAAATAAAGAA
CAAGAATGGAGTAATACTAATGATGTTTACATGGTAGAGTATCAAATGGATTCTGTGAGC
ATCATAGAGAAAGTTTCTTGAGACTAACCAGTGAAGTCTAGCTCCTCTTCCATTTTGGTGAC
TCCGAGGAATGTTACACAAGTCTTAGTTAGTTAGTAAGAGGTACAATTTAGCAGCAGGTAAC
AGCAATGTGTTAACTAAAAACCTAATTGAGTAATGCAGTTTTGATTAAATATTAGCTTTT
TGGTAATTCAGGAATGTCGACACCAAGGG
>G2383 Amino Acid Sequence (conserved domain in AA coordinates:89-149)
MFPSFITHIQSPNSHHYSSPSFPFSSDFLESFDESFLINQFLQQQDVAANVVEPWKF
CKKLELKKKNEKCVDGSTSQEVQWRRTVKKRDRHSKICTAQGPRDRMRSLQIARKFFD
LQDMLGFDKASKTIEWLFSKSKTSIKQLKERVAASEGGGKDEHLQVDEKEKDETLKLRVS
KRRTKTMESSEFKTKESRERARKRERERTMAKMKMRLFETSETISDPHQETREIKITNGVQ
LLEKENKEQEWSTNDVHMVEYQMDSVSIIKFLGLTSDSSSSSIFGDSSECYTSLSSVR
GTISAAGNSNVLTKNPN*
>G571 (326..1708)
TAGCCGACCTCTCTTCTCTTCTGAAAAAACACCAAAGGAGCTTTAAATGCTCCGTTA
CATAATCTCTATCTCTTCCAGAATATAGAGAAAGGAAAATAATATACAAGAATTAATA
GAAGGTATATCATCATCTCTCTAGCTAGTGATCAAAGCACCGTCATCATCATATATC
ATCAGCTTGCCCTCAGAGGAGAAGACCAACATAAGAGAGATCGAAGATCAAATCTATCTC
TCTTCATCATCTTCTGCTGTTACTATCATATCACACGCTCTCTCAAACATCATCTATAT
ATAGACTTCTCTTCATCATCAATGCAAGGTCATCACCAGAATCATCATCAACACTT
ATCATCATCTCCGCCACGCTCTTCCCATGGAAGTTCATGAACAAAGATGGGTATGATAT
TGGAGAGATAGACCCATCACTCTTCTCTATCTTGATGGACAAGGACATCATGATCTCC
ATCAACTGCTCTCTCTCTCTTACATCATCATCACAACTCAGAATTTGGCGATGAGACC
TCCAACATCGACGCTCAACATCTTCCATCTCAGCCTATGCACATAGAGCCACCTCTTC
TTCTACACACAATACCGATAATACAAGATTAGTTCGGGCTGCTCAACCTAGTGGTTCCAC
TCGACCAGCTTCTGACCCGCTCCATGGACTTGACCAATCATTCTCAGTTTCATCAACCTCC
TCAAGGTTCTAAATCCATCAAGAAGGAAGGGAACCGCAAGGGTCTTGCTCATCGGACCA
TGACATACCTAAATCGTCAGACCCTAAACATTGAGAAGACTAGCACAAAACAGAGAAGC
AGCAAGAAAAAGCAGATTACGTAAAAAGGCTTATGTTTCAAGCACTCGAGTCATGTAGGAT
CAAACGTACCCAACTAGAACAAAGAGATTCAACGGGGCCAGATCCCAAGGCGTATTCTTTGG
AGGGTCTCTTATAGGAGGAGATCAACAGCAAGGTGGACTACCATTTGGCCCTGGCAACAT
CAGCTCTGAAGCAGCGGTGTTGATATGGAATATGCGAGGTGGCTGGAGGAGCAGCAGAG
GCTATTAAACGAACTAAGGGTGGCAACACAAGAACAATGTCGAGAACGAGCTTAGGAT
GTTTGTGGACACATGTTTAGCTCATTATGACCATTGATTAACTCAAGGCTATGGTCGC
TAAGACCGATGTCTTCCACCTCATTTCTGGAGCATGGAAGTCCAGCTGAACGTTGCTT
CTTGTGGATGGGTGTTTTCCGTCCATCGGAGATCATTAAGGTGATTGTGAACCAGATAGA
ACCATTGACGGAGCAACAGATAGTTGGGATATGTGGGCTGCAACAGTCCACACAAGAGGC
CGAGGAGGCTCTCTCGAAGGCTCGAGGCGTTGAATCAATCACTTCCGATAGCATTGT
CTCTGACTCCCTCCCGCTGCTCCGCACCACTTCTCTCTCATCTATCCAATTTATGTC
ACACATGTCTTAGCTCTCAACAAGCTCTCTGCTCTCGAGGGCTTCGTTCTCCAGGCGGA
TAATTTGAGGCACCAAACGATCCATAGGCTGAACCAATTGTTGACGACCCGTCAAGAAGC
ACGGTGTCTTTAGCGGTTGCGGAGTACTTCCACCGTCTTCAAGCTCTAAGTTCTCTCTG
GCTAGCCCGTCTCGGCAAGATGGATAATACTAAAACAACTGATGAAGGAAACCAAAAAC
AAAAACAAGAGAATAGGTTGATTAGTTAGCCGCGCAGCTTGACCTCTTTATCATATATATC
GTCTCTCTACTCAAATACAGTGCAATTAGGGAAAATGTTTGGCTTCTTTTGGTATATG

ATTCTTACTATTATGTTTTTAATCAAGA

>G571 Amino Acid Sequence (domain in AA coordinates: 160-220)
MQGHHQNHQHQLSSSSATSSHGFMNKDGYDIGEIDPSLFLYLDGQGHDPSTAPSPPLH
HHHTTQNLAMRPPTSTLNIPSPQPMHIEPPPSSTHNTDNLRLVPAAPSGSTRPASDPSM
DLTNHSQFHQPPQGSKSIIKEGNRKLGLASSDHDIPKSSDPKTLRRLAQNREAAKRSRLRK
KAYVQQLESCRIKLTQLEQEIQRARSQGVFFGSLIGGDQQQGGPLPGPGNISSEAAVFD
MEYARWLEEQRLLNELRVATQEHLSENELRMFVDTCLAHYDHLINLKAMVAKTDVFLHI
SGAWKTPAERCFLWMGGFRPSEIIKVIVNQIEPLTEQQIVGICGLQQSTQEAEEALSQGL
EALNQSLSDSIVSDSLPPASAPLPPHLSNFMHMSLALNKLKSALEGFVLQADNLRHQTIH
RLNQLLTTRQEARCLLAVAEYFHLRLQALSSWLARPRQDG*

>G636 (6..1814)

CGATGATGCAACTGGGTGGTGGTACTCCGACCACTACAGCGGGGGCTACAACCGTCACAA
CTGCTACAGCACCACCGCCACAATCAAACAACAACGATTACAGCGGCAACAGAAGCAGCGG
CAGCAGCGGTGGGGCGTTTGTAGGTGTGCGAAGAGATGCACGACCGTGGGTTTGGAGGAA
ATCGTTGGCCCGCGCAGGAAACGCTAGCGTTGTTGAAAATACGATCTGACATGGGAATAG
CGTTTCGAGACGCTAGCGTTAAAGGTCCCTTATGGGAAGAGGTTTCTAGGAAAATGGCGG
AGCATGGTTACATAAGAAACGCAAGAAATGCAAAGAGAAATTCGAGAACGTTTACAAAT
ACCAAAACGAACCAAGAAAGTTCGTACCGGAAAATCCGAGGCAAACTTATCGCTTCT
TTGATCAATTAGAAGCTCTCGAGTCTCAATCTACAACCTCACTCCACCATCATCAACAAC
AAACGCCTCTTCGACCACAGCAAAACAACAACAACAACAACAACAACAACAACAGCT
CCATATTTTCAACTCCTCCTCCGTTAAGCAGCTTATGCCGACGCTTCTCTTTCATCAA
TTCCTCCGTATCTCAGCAGATTAATGTACCTTCGTTTCCAAACATCTCCGGTGATTTTC
TATCGGATAATTCTACATCGTCTTCGTCTTCTTATTCGACTTCTTCTGACATGGAGATGG
GTGGTGGAACGCGACTACAAGGAAGAAAGAGGAAATGGAAGGTGTTTTCGAGC
GGTTGATGAAACAAGTAGTTGATAAACAAGAGAGCTTCAACGCACATTCTTGAAGCTG
TTGAAAAGCGAGAACACAAGAGATTGGTTAGAGAAGAGTCTTGAGAGATTCAAGAGATTG
CCAGAATCAACCGCAGCAGCAGATCTTAGCTCAAGAACGCTTATGTCCGCTGCAAAAG
ACGCTGCTGTTATGGCCTTTCTTCAAAAACGCTGAGAGAAACAACCGAATCAGCCACAAC
CGCAGCTCAGCCGCAACAAGTTCGACCATCAATGCAGCTTAATAACAACAATCAGCAGC
AACCGCTCAACGGTCTCCTCCACCGCAACCTCCTGCTCCGCTTCCGAGCCAATTCAGG
CGGTTGTGTCGAGCTTAGACACAACGAAAACGCACAATCGTGGTGATCAGAATATGACTC
CTGCAGCTTCAGCGAGCTCGTCCGCGTGGCCGAAAGTGAGATAGAAGCATTTGATAAAGC
TGAGGACGAATCTTGATTCGAAATATCAAGAAAACGGACCAAAAGGACCATTGTGGGAAG
AGATATCAGCGGAATGAGAAGGTTAGGATTCAACAGGAACCTCAAAGAGATGCAAAGAGA
AATGGGAAAACATAACAATACTTCAAGAAAGTCAAAGAGAGCAACAAGAAACGTCCTCCG
AAGATTCCAAGACTTGCCCTTACTTTCACAGCTTGATGCTTTATATAGAGAGAGGAACA
AATTCACAGCAACAACAACATTGCAGCTTCTTCTTCTATCTTCCGGTCTTGTAAACCGG
ATAATTCTGTTCCCTTGATGGTCCAACAGAGCAGCAATGGCCTCCGGCTGTAACGACTG
CGACAACCTACTCCCGCAGCGGCTCAGCCTGATCAGCAATCTCAGCCGTCGGAGCAGAACT
TTGATGATGAAGAAGGTACAGATGAAGAGTACGACGATGAAGATGAGGAAGAGGAGAATG
AAGAAGAGGAAGGAGGTGAGTTCGAGCTTGTGCCTAGCAATAACAACAACAAGACGA
CGAATAATCTGTAATGATGATGATTGAGTTCGAACCGGTTTGGTGGTGAAAGATTAGTA
ATCTTTTTTTAAGTTTTGATACAGAACATGAGAATTTAAATATTGGAGGGTTT

>G636 Amino Acid Sequence (domain in AA coordinates: 55-145, 405-498)

MQLGGGTPTTTAAATTVTATAPPPQSNNDSSAATEAAAAVGAFAVSEEMHNRGFGGGR
WPRQETLALLKIRSDMGIAFRDASVKGPLWEEVSRKMAEHGYIRNAKKCKEKFENVYKYH
KRTKEGRTGKSEKTYRFFDQLEALESQSTSLHHHQQTPLRPQQNNNNNNNNNNSS I
FSTPPPVTVMPTLPSSSIPPYTQQINVPSPFNISGDFLSDNSTSSSSSYSTSSDMMGG
GTATTRKKRKRKWKVFFERLMKQVVDKQELQRTFLEAVEKREHKRLVREESWRVQEIR
INREHEILAQERSMSAAKDAAVMAFLQKLSEKQPNQPQPQPQVPRPSMQLNNNNNQQP
PQRSPPPPAPLPPQPIQAVVSTLDTTKTHNRGDQNMTPAASASSSRWPKVEIEALIKLR
TNLDSKYQENGPKPLWEEISAGMRRLGFNRSKRCKEKENINKYFKVKESNKKRPED
SKTCPYFHLQDALYRERNKFHSNNNNIAASSSSGLVKPDNSVPLMVQPEQQWPPAVTTAT
TTPAAAQPDQSQPSEQNFDEEGTDEEYDDEDEEEENBEEEGGEFELVPSNNNNNNKTTN
NL*

>G878 (197..1738)

CAAAAAAATCTCTCCCATTAAGAACTGCCCAAAGAAATATTTTATACAAAATGAAAGA

GAGAAACACGACACGAATTTTGTATAATTAAGATTACACAAAAAAGTGTAGAAAGAG
AAATATCTTCTTCTTTTCTGTGTGAGTTGGGTTTGTAAAGTTTATCCTTTTGTTC
TCAAAATCAAGAATCGATGGCGGAGAAGGAAGAAAAAGAACCATCGAAGTTAAAATCATC
CACCGGAGTTTACGGCCAACGATTTCACTACCTCCTCGACCGTTTGGTGAAATGTTTTT
TAGCGGTGGCGTTGGATTTAGTCTTGACCAATGACTCTCGTCTCAAATTTATTCTCTGA
TCCTGATGAGTTCAAGTCTTCTCTCAGCTTTAGCTGGAGCTATGGCTTCTCCGGCGGC
AGCTGCTGTTGCCGCCGCTGCTGTGGTTGCTACTGCTCATCATCAGACACCTGTGAGCTC
TGTCGGTGATGGCGGTGGAAGCGGTGGTGATGTTGACCCGAGGTTTAAAGCAGAGTAGACC
AACCGGATTGATGATAACTCAACCACCGGGGATGTTTACTGTACCGCCGGGGTTAAGTCC
GGCTACTCTTTTGGATTCTCCGAGCTTCTTGGTCTTTTTCACCTCTTCAGGGAACATT
TGGTATGACACATCAACAAGCTTTAGCACAAGTCACTGCACAAGCAGTTCAAGGCAATAA
TGTTTCATATGCAGCAATCACAACAATCTGAATATCCTTCTTACACAACAACAACA
ACAACAACAACAAGCTTCAATGACTGAGATTCCATCATTTTCTTCTGCACCTAGGTCTCA
GATTCGAGCCTCGGTTCAAGAAACATCGCAGGGTCAGAGAGAGACTTCGGAATATCTGT
CTTTGAGCATCGGTACAGCCTCAAAATGCTGACAAACCAGCTGATGATGGATACAACTG
GCGGAAATATGGGCAGAAAGCAAGTGAAGGGGAGCGATTTTCTCGAGTTATTACAAATG
TACGCATCCAGCTTGTCTGTCAAGAAGAAAGTGGAGAGGTCACCTCGATGGACAAGTAAC
GGAAATCATCTACAAGGGTCAACACAATCATGAGCTTCTCAAAAGCGCGGTAACAATAA
CGGGAGTTGTAAAGTTCTGATATTGCAAAATCAGTTTCAACAAGTAATAGCAGTCTCAA
CAAGAGTAAGAGGGACCAAGGAAACAAGCAAGTTACAACAACAGAGCAGATGTCTGAAGC
AAGTGATAGCGAGGAGTTGGGAATGCAGAGACTAGTGTGGGAGAAAGACATGAGGATGA
GCCTGATCCCAAGCGAAGAAATACAGAAGTTCCGGGTTTCAAGACCAAGTTGCTTCATCGCA
TAGAAGTGTGACAGAGCTTAGGATTATTGTCCAAACGACGAGTGAAGTTGACCTCTTAGA
TGATGGATATAGGTGGCGCAAGTATGGTCAGAAAAGTAGTCAAAGGAAATCCTTATCCGAG
GAGCTACTATAAGTGTACAACACCAGATTGCGGAGTAAGGAAACATGTAGAGAGAGCAGC
AACTGACCCAAAAGCTGTTGTAACAACATATGAAGGTAAACATAACCATGATGTTCCAGC
TGCTAGAACCAGCAGCCATCAGTTAAGACCAAAACAATCAACACAACACCTCAACGGTTAA
CTTCAATCATCAACAGCCTGTTGCACGTTTAAAGGCTTAAAGAAGAGCAAACTCACTTGACA
GAGAAGAAGAATACGACGGCGCTTGAGCTTTTGTGAGTTTAAATGAATCTTCTTTTGGTT
AATGAACCTGTTTTTGTGCTTCAAAAACACCACAGGTTTCTCTGGACAGAATCTCTGATA
TTACAGTTTCAAAAGGTATGTTCTTTTATTTCATGTTGGAATCTTCTGTGAATCTTAAG
AAGCTTTAGGAGGTAAATGTAAAAAACAGATTCAAAGTTATGCCCTTATGTGAATCTTT
TGATACATGGGATAAACAATTTACAGGTATCCTTTTTTGTCTTGTGTAACAAAAA
AAAA

>G878 Amino Acid Sequence (domain in AA coordinates:250-305, 415-475)

MAEKKEEPSKLSSTGVSRPTISLPPRPFGEFFSGGVGFSPGPMTLVSNLFSDPDEFK
SFSQLLAGAMASPAAAVAAAAVVATAHHQTPVSSVGDGGSGGDVDRFKQSRPTGLMI
TQPPGMFTVPPLSPATLLDSPSFFGLFSPLQGTFGMTHQQALAQVTAQAVQGNVHMQQ
SQQSEYPSSTQQQQQQQASLTEIPSFSSAPRSQIRASVQETSQGGQRETSEISVFEHRS
QPQNADKPADDGYNWRKYGQKQVKGSDFPKSYKCTHPACPVKKKVERSLDGQVTEIYK
GQHNHELPQKRGNNNGSKSSDIANQFQTSNSSLNKSQRDQETSQVTTTEQMSEASDSEE
VGNAETSVGERHEDEPDPRRNTTEVRVSEPVASSHRTVTEPRIIVQTTSEVDLLDDGYRW
RKYGQKVVKGPNYPRSYKCTTPDCGVRKHVERAATDPKAVVTTYEGKHNHDVPAARTSS
HQLRPNNQHNTSTVNFNHQPPVARLRLKEEQIT*

>G1134 (61..849)

TAAAGAAAGAGAAAAAAGCTTTTCGTAGTGTCTATTGAAACCAGAGAAAAGCCAAAGGGG
ATGCAACCAACATCCGTCGGTAGTAGCGCGGTTGGTGACGACGAGGAGGAGGAGGAGGA
GGAGGAGGGCTAAGTAGAAGTGGACTATCTCGGATCCGTTTCAGCTCCAGCGACTTGGCTT
GAAGCTTTACTTTGAGGAAGATGAAGAAGAGTCTTTGAAACCTAATCTTGGTCTCACCGAT
TTGCTTACCAGGAACTCGAACGATTTACCGACAAGTCGCGGCTCGTTTCAGATTCCCAGATT
CCTGTTGAGCAAGGGTTGATCAACAAGGTGGGTTTACCGACAGAATAGTACTCCGGCG
GATTTTCTTGTAGTGGTTCTGATGGATTATCCAAAGCTTTGGGATTGAGGCGAATTACGAT
TACTTATCGGGGAATATCGATGTTTCTCCGGGAAGTAAGCGGTCTAGAGAAATGGAAGCA
CTCTTCTCTCTCTGAGTTTACTTCTCAAATGAAAGGAGAGCAAAGCAGCGGTCAAGTT
CCTACCGAGTATCAAGCATGTCGATATGAACATGGAGAACCCTTATGGAGGACTCTGTT
GCTTTTAGGGTTCCGGCTAAACGTGGTTGCGCAACTCATCCCCGAGCATTGCCGAGAGG
GTACGAAGGACGCGGATTAGTGATCGGATAAGGAAGCTACAAGAGCTTGACCTAACATG

GACGAAGCAAAACCAACACTGCAGACATGTTAGAGAAGACAGTAGTAATACCTGTAAGTTCTT
CAAAGGCAGATCCAGGAGTTAAACAGAAAGACAGAAGAGGTGCACATGCATACCTAAGGAA
GAACAATAAGGTTTGCTCCTGATTTGTTTATATTTGCTTAACGGCAATGATCTGATCGA
AAAATTCGAAAGATGATCTTAGCTTGAATTTAGATGGATGTCATGTTGAAAAGTATATTA
TTTGATAAATGGATGTAGGTGTAATATAAAATTTTTGTACAATAATGAAGAAAGTTAAAA
AGAATTAATGAAAAACATATATTTTATGATATAAAAAA

>G1134 Amino Acid Sequence (domain in AA coordinates: 198-247)
MQPTSVGSSGGGDDGGGRGGGGGLSRSLRIRAPATWLEALLEEDEEESLKNPLGLTD
LLTGNSNDLPTSRGSFEFPIVEQGLYQQGGFHRQNSTPADFLSGSDGFIQSFGIQANYD
YLSGNIDVSPGSKRSREMEALFSSPEFTSQMKGEQSSGVPTGVSSMSDMNMENLMEDSV
AFRVRAKRGCAATHPRISIAERVRRTRISDRIRKLQELVPMNDKQNTADMLEEAVEYVKVL
QRQIQELTEEQKRCTCIPKEEQ*

>G1008 (89..973)
GCCTTTTTGACTCTTCTTTCTCTCTTCTACTTTTTTTTCAGGCTCTCTCTCTATATCTCTA
TCTTCTTCTCCGGTTAACTAAAAAGAGAAATGAAAAGCCGAGTGAGAAAAATCCAAGTACAC
GGTTCACCGGAAAATCACATCCACACCGTTTCGACGGTTTCCCGAAGATTGTCAAAATCAT
AGTCACTGACCCATGCGCTACTGATTCTTCCAGCGATGAGGAAAACGACACAAATCTGT
TGCTCCGAGGGTGAAACGTTATGTGGATGAGATCAGGTTCTGTGACGAAGATGACGAACC
TAAACCGGCGGAGGAAAGCGAAGAAAAAGTCCCCGGCGGCTGCGGCGGAGAACGGTGGAGA
TTTGGTAAGTCTGTGGTGAAGTATAGAGGAGTGAGACAACGACCTTGGGGAAAATTTGC
GGCGGAGATTCTGTGACTCTTGAGTCTAGACTAGACTCTGGCTTGGGACTTTTGGCAGCGC
GGAGGAAGCTGCTATAGGTTACGATAGAGCCGCGATTGGAATCAAAGGTCATAACGCTCA
GACGAATTTTCTACTCTCTCTCTAGTCCGACGACTGAGGTGTTACCGGAAACTCCGGT
GATTGACCTTGAAACTGTCTCTGGTTGTGATTGGGCGAGGGAATCGCAAATCAGTCTGTG
TTCPCGACTTCTGTTCTCCGGTTTAGTCACAACGACGAAACAGAGTACAGAACAGAGCC
AACGGAAGAACAAAATCCGTTTCTTGCCCTGATTTGTTTCGCTCCGGAGATTATTTTTG
GGATTCCGAAATTACCCCTGACCTTTGTTTCTCGACGAATTCACCACTCTGTTGTTACC
AAATCATCAACAACAACAACACAGCTGTGTGATAAGGATACGAATCTGTCTGATGATTTTCC
GTGGGAGTGATCGGAGATTCTAGCTCATGGGATGTTGATGAGTTTTTCCAAGATCATT
GTTGGATAAGTAATTTGATGAGTTCTTCCCAGAAATTTTTCTGGGTTTCTCTTTTTGGTT
GTGTGAGTGAGATGAGTGGTTTGATGACAACGACGGGGATGAATCTTAGCCGTCGGTTTT
CCATTTCTGTGGACGGCTCCGATCAGCGGAAGAGCGCAACGGAGTTTTTATTTATCTGTT
TGAGAATTTTATAATTTAATTTGCGAGTAAATATAGTAATTAGTGTTAAGATTGTGAGAG
TTTAAGTTAATTAGGGAGGGGTTTGAATATTGGGGATTTTGGGAGGTTTTTGTGTTGGTT
TCTCTCAAAGTCTGTCTACTATGTCAGGAAGCAGGATATAAGACCGTATATATATTTATTA
TTAATATTGATAAAAGTAAAAA

>G1008 Amino Acid Sequence (domain in AA coordinates: 96-163)
MKSVRVRSKYTVHRKITSTPFDGFPKIVKIIIVDPCATDSSSDEENDNKSVAAPRVKRYVD
EIRFCDEDDPKPARKAKKSPAAAAENGDDLKSVVVKYRGVQRQRPWGKFAEIRDPSSR
TRLWLGTAFATAEEAAIGYDRAAIRIKGHNAQTNFLTPPPSPTEVLPEPVIDLETVSGC
DSARESQISLCSPTSVNLRFSHNDETEYRTEPTEEQNPFPLPDLFRSGDYFDWSEITPDPL
FLDFHFQSLPLNNINMNTVCDKDTNLSDSFPLGVIGDFSSWDVDEFFQDHLDDK*

>G1020 (132..689)
CTGTTCAACAAGAAAGCTCCCCAAAAGGAGCGTTGCTTTACTCTCTCTATAAAAAGAAGCTC
TTCTACTTCTTCTCGTTACCAAAAACCTCTTTCACCGATCTTCTCGTTCCATTCTTCTTC
CTAATTACACCATGCCAACATACCATGGGTTTGAAACCCGACCGGTTGCTCCAACGA
ACCCGACTCATCATGAGAGTAATGCTGCCAAAGAGATTGCTTACAGAGGCGTTAGGAAAC
GTCCATGGGGAAGAFAACGCCGCTGAGATCCGAGATCCGGTTAAGAAAACCTCGATCTGGC
TCGGTACGTTTCGACACCGCTCAGCAGCGCGCGTGTCTACGACGACGCCGCGCTGACT
TCTCGTGTTGTAAGGTTAAGACCAATTTCCGGTGTGTTCTGTTGGTAGTAGTCTTACTCAGA
GTAGCACCGCTCGTCTGACTCTCCCACGGCGGCACGGTTTATAACACCTCCGCACCTCGAGC
TCAGCTTAGGCGGCGCGCGCGTGTGTCGTAAGATCCCGCTTGTGCATCCGGTTTACT
ACTATAACATGGCGACGTATCCAAAGATGACGACGTGTGGTGTCCAGAGCGAGTCTGAAA
CGTCGTCGGTCTGTTGATTTCAAGGTTGGAGCTGGGAAAGATATCTCCGCGGTTAGATCTGG
ATCTTAACCTTAGCTCTCCGCGGAATAGGCGGTGAGTTTTTTTTTCTTATGTCGTTTC
TTTAGACAAAAAATAACGTTTCTTTTCTGCTTAAGAAAAAATATTATCCG
TTTTTTAGAAAGAAAAA

>G1020 Amino Acid Sequence (domain in AA coordinates:28-95)
MPNITMGLKPDVPAPNTNPHHESNAAKEIRYRGVVRKRPWGRYAAEIRDPVKTRVWLGT
DTAQQAARAYDAAARDFRGVKAKTNFGVIVGSSPTQSSTVVDSPTAARFITPPHLELSLG
GGGACRRKIPLVHPVYYNMATYPKMTTCGVQSESETSSVDFEGGAGKISPPLDLNL
APPAE*

>G1023 (252..1250)

TCGTCTTCTTAATCGCTTTCTGCTCTGTTTTCTCGTTCATCAAGCTACATCTACTAGCT
CTCTCAGTGATGTATTTCTCACAGTTTCATCGATTTCATGCGTTTAAAGACCTAAAAGGA
CTTGTCTGAGGTAAGGACTTTTCTGTCTTGAGAGAGTTCATTTTGAGGCTTTTCTG
GGAATTTTGAGAGGTTTTTAGGGTTTAAAGGGGTTTGGTTTTGAATTTTCGCACACCAAG
TGTTTCGATAAATGGCTGAACGAAAGAAACGCTCTTCTATTCAAACCAATAAACCAACA
AAAAACCCATGAAGAAGAAACCTTTTTCAGCTAAATCACCTCCCAGGTTTATCTGAAGATT
TGAAGACTATGAGAAAACCTCGTTTCGTTGTGAATGATCCTTACGCTACTGACTACTCAT
CAAGCGAAGAAGAAGAAAGGAGTCAGAGAAGGAAACGTTATGTCTGTGAGATCGATCTTC
CTTTCGCTCAAGCTGCTACTCAAGCAGAATCTGAAAGCTCATATTGTCAGGAGAGTAACA
ATAATGGTGTAGCAAGACTAAAATCTCAGCTGTAGCAAAAAGGTTTACGCAGCAAG
CATCTCCGGTCGTTGGACGTTCTTCTACTACTGTCTCGAAGCCTGTTGGTGTAGGCAGA
GGAAATGGGGTAAATGGGGTCTGAGATTAGACATCCATCACCAAGTAAGAACTTGGT
TGGGTACTTACGAGACGCTGAACAAGCAGCTGATGCTTATGCTACCAAGAAGCTTGAGT
TTGATGCTCTGGGCTGCAGCACTTCTGCTGCTTCTCTGTTTTGTCAAATGAGTCTGGTT
CTATGATCTCAGCCTCAGGGTCAAGCATTGATCTTGACAAGAAGCTAGTTGATTGCACTC
TTGATCAACAAGCTGGTGAATCGAAGAAAGCGAGTTTTGATTTGACTTTGCAGATCTAC
AGATTCTGAAATGGGTTGCTTCATTGATGACTCATTCATCCCAATGCTTGTGAGCTTG
ATTTTCTCTTAACAGAAGAGAACAACAACCAATGTTGGATGATTACTGTGGCATAGATG
ATCTGGACATCATTGGTCTTGAATGTGACGGTCCAAGCGAATCTCCAGACTATGATTTCT
CAGATGTGAGATCGATCTTGGTCTCATTTGAACCAACCATGACAAGTATGCTTTCGTTG
ATCATATCGCAACAACACTACTCCCACTCTCTTAATATCGCGTGCCCATAGTTTGCAGC
TAGGTGTTATTATTAGCTATAGGAGCAACGTAAGGCTCGTTGTTACTCGGTTTTGTCT
TAAGTTATTAAAGTATAGCAGAGGCAGTTAATCTCAAGGGAAGCAAAAACCTAAAGATA
GAAGCAGATGCAGTTTTGTGTGTTGGTGTACTAAAGAAAGTTTTGTTGACATAATGGTT
TTGATGTTGTGGAGAAGATAGAGAGGTGTGATCGAAATTGTAATCTCAGGTGGTTTTTT
TTGAAGGCAATTGTTCTCATTAGGGTTTTTTTCTATATGAGGATTGTCTTTGAAAAGC
CTTTAGATGTTTTCTAATTCGTAAGCTCTCTCAATCTTTGTAAGTTTTGCCTGTTGAGTT
ATTGATACATATGTGAGACCTACTTTATTTGTTTTGTGCTACATACATTGTTGATGGTT
CGTCAAAAAAAA

>G1023 Amino Acid Sequence (conserved domain in AA coordinates:128-195)
MAERKRSSIQTNPKNKPKMKPFQLNHLPGLESLDKTMRKLRVFNNDPYATDYSSSEE
EERSQRRKRYVCEIDLFPFAQAATQAESESSYQESNNNGVSKTKISACSKVLRKASPV
VGRSSTTVSKPGVVRQRKWKWAAEIRHPITKVRTWLGTYETLEQAADAYATKKLEFDAL
AAATSAASSVLSNESGSMISAGSSIDLKLVDSLDQQAGESKKASFDFDLQIPE
MGCFIGDSFIPNACELDFLLTEENNNQMLDDYCGIDDLDIIGLECDGPSELDPDYDFSDVE
IDLGLIGTTIDKYAFVDHIATTTPTPLNIACP*

>G1053 (38..538)

GAAACTCTTACATACTCATATAAACCAAACTAAAACCATGATTCCGGCAGAAATCAACGG
ATATTCCAATATCTATCACCGGAATACAACGTAATAAACATGCCTTTCATCTCCAACTC
TTCTTAAACTACCTAAACGATTTGATCATCAACAACAACACTATTCTCATCATCCAA
CAGTCAAGATCTCATGATAAGCAACAACCTCAACTCCGACGAAGATCATCATCAAGCAT
CATGGTACTCGACGAGAGGAAACAGAGAAGGATGCTTTCGAACAGAGAATCTGCAAGGAG
GTCAAGGATGAGGAAACAGAGACATCTTGATGAACTCTGGTCTCAGGTAATAAGGCTTCG
CAACGAGAACAACCTGTCTTATCGATAAGCTGAACCGGTATCGGAGACTCAAAATTGTGT
ATTGAAGGAGAACCTTAACTCAAAGAAGAAGCTTCTGATCTCCGACAGCTTGTGTGA
ACTGAAATCTAACAAGAACAACAACATAGTTTTCCAAGAGAGTTTGAAGATAATTAGTA
TTACTCAA

>G1053 Amino Acid Sequence (domain in AA coordinates: 74-120)
MIPAEINGYFQYLSPEYNVINMPSSPTSSLNYLNDLIINNYYSSSSNSQDLMISNNSTS
DEDHQSIMVLDERKQRRMLSNRESARRSRMRKQRHLDELWSQVIRLRNENCLIDKLR
VSETQNCVLKENSCLKKEASDLRQLVCELKSNKNWNSFPREFEDN*

>G1137 (202..1248)

TACTTCAGACTTCTACTCAAACCAAGTCACGTAGTTGGTTGGTGACATTTTCGCTGCATTTT
TCAATCTGTGATTGTTTTTCGTTTCGTCTTTCTTTTACTATTTCCTCGAAAAGGACACAAG
AAGTATTGCATTCACTCAGTTGAGCAACTTAACAATCGTGTGTACTTTTTGAAGTTCCC
TTGAGCTAAACTGCTAAGAGCATGCCTCTGGATAAGAGGCAACGGGATTTGCCTCTGGGC
TTAAGTCCTCAAGCTTGCTTCAAGGATATAGTAGGTGGTCTGTCTTCCTAGAATTCCT
CTCCCTGAGCTTGGGAACTATATGCAGCTAAGCTTCAGGCTCGCTGTTTGCAGCCACCA
CCATTCCAGTCTTTGCTGTGCAGTCATGATAAGGAGTCTTATGGAAAAAGATTCTCACGG
TCTGACATGCGGTCTTGGTGCGCTGTCTACTACTACTACTTCCACTTGGAGCATT
GAGTCTTCTCAGAAAAGACTTTTGATATTTCGATCAGTCAGGAGACCAGACTCGTCTATTA
CAATGTCCATTTCTCTACGGTTCCATCTCATGCGGTGCAGAACCAAGTGAAGTCTCT
GAGTTACAAGGTATAGAGAAAGCTTTCAAAGAAGATGGTGAAGAGTTTACAAGAGTGAT
GGAACAGAGTCAGAAATGCATGAAGACACTGAGGAGATCAATGCATTGCTATATTCAGAT
GATGATTATGATGATGATTGCGAGAGTGATGATGAAGTAATGAGCACTGGTCACTCTCCT
TATCCAAATGAAGGAGTTTGCAACAAAAGGGAATTAGAAGAAATCGATGGTCTTGTA
AGGCAGAACTACTGGATAAGGTCAACAACATCAGCGACTTATCATCACTTGTGGGCACT
GAGAGCTCCACACAACCTCAATGGATCTTCTTTCTTAAGGACAAAAGCTCCCTGAATCA
AAAACCATATCGACCAAGAGGACACTGGTTCTGGTCTGAGCAACGAGCAGTCGAAGAAA
GACAAGATCCGACAGCTCTGAAAATACTCGAGAGCGTAGTCCCTGGTGAAAAGGAAAC
GAAGCGCTCTTACTTCTGGACGAAGCAATTGATTACCTAAAGTTGCTGAAACGAGACTTA
ATCTCCACAGAGGTTAAGAACCAAGCTCCACCACTACAAGTCACCAATCTTGTGCTT
AAAGAGACAACATGGGGACAAGAAATCTGCAGACAGATAAGGCGTGAAAGATTCTGACG
AGTTAAAACGTGTGAAGTGGGTTTTTGGGTACGTATCCTTGCACCAGCTTT

>G1137 Amino Acid Sequence (domain in AA coordinates: 264-314)

MPLDKRQRLDPLGLSPQACFKDIVGRSVLPRIPLPELGKLYAAKLQARCLQPPPFQSLLC
SHDKESYGKRFSRSDMRSGWCAAAATTTTFLGALESSQKRLILFDQSGDQTRLLQCPFLR
FP SHAAPVKLSELQIEKAFKEDGEEFHKSDGTESEMHEDTEEINALLYSDDDYDDDC
ESDDEVMSHGSPYPNEGVCNKRELEIDGPCKRQKLLDKVMNISDLSSLVGTESSTQLN
GSSFLKDKKLPESTKISTKEDTGSGLSNEQSKDKIRITALKILESVPVPAKAGNEALLLLD
EAIDYLLKLLKRLISTEVKNQSSSTTHKSPILLKETTGWTRNLQTDKA*

>G1181 (113..1012)

CTCGATCTTTTAACCCCATTTATTACATATTACTCCTTCCTACATTATTCTTCTTCTGCT
TTCGTGACTTTTCAGGGGACACTTTTGTTTTATAACTTACGCTTAAATCCTATGAATTC
GCCGCCGTTTGACGCAATGATTACCGGAGAATCATCGTCACAAAGATCTATCCCAACGCC
GTTTCTCACAAAACGTTTAACTCGTTGAAGATAGTTCCATCGACGATGTTATCTCATG
GAACGAAGATGTTCTCTTTTCATCGTATGGAATCCGACAGATTTTCGCTAAAGATTTGCT
TCTTAAACACTTCAAACACAACAATTTCTCTAGTTTCGTTTCGTCAGCTCAACACTTACGG
ATTCAAAAAGTTGTACCGGATCGATGGGAGTTTCAAACGATTTCTTTAAGAGAGGAGA
AAAACGTCTTCTCCGTGAGATCCAACGTCGGAATAACACGACGCATCAAACAGTTGT
TGCTCCTTTCGTCGGAACAACGAAACAGACGATGGTTGTATCACCGTCAAATTCGGGGA
AGATAATAATAATAATCAGGTGATGTCTTCGTCTCCGTCTCGTGGTATTGTCATCAAAC
GAAGACGACTGGGAATGGTGGTTTATCAGTGGAGTTATTGGAAGAGAACGAGAAGCTTCG
GAGTCAAAACATTAGCTAAACCGTGAGCTTACTCAGATGAAATCTATCTGCGATAATAT
CTATAGTCTCATGTCCAATTACGTCGGATCTCAGCCCACTGATCGGAGTTATTCTCCCGG
AGGTAGTAGTAGTCAACCGATGGAGTTTTTACCGGCGAAGCGGTTTTTCGGAGATGGAGAT
TGAAGAAGAAGAAGACGAGTCCGAGGTTGTTTGGTGTTCGATTGGGTTAAAACGGAC
GAGAAGTGAAGGTGTTCAAGTGAAGACGACGCGGTTGGTGGGAAAATTCCGATGAGGA
GACGCCGTGGTTGAGACATTATAATCGAACCAATCAGAGAGTTGTAATTAACGAAC
GGTTTAGATTTGTGGTGTAGATATGTGCGCGAAGTAGACGATTACAGCTTTTAAAGACAA
GCAGAGCACGTGTCCATCTGTTTCAAGAAGTTCTGCAATCTTGACTTCTTCTTTTAAAC
ACTTTGTGTTTTTTTATTTAATTAATAACAATAAATGTTCTTTTTCAGTTTGTGTTTT
TTCAAAAATAGTTCCGCTGTTTCTAGACTTTCCTTTTTT

>G1181 Amino Acid Sequence (domain in AA coordinates: 24-114)

MNSPPVDAMITGESSQRSIPTPLTKTFNLVEDSSIDDVISWNEDGSSFIWNPTDFAK
DLLPKHFKHNNFSSFVRQLNTYGFKKVVPDRWEFSNDFFRGEKRLLEIQRKITTTHQ
TVVAPSSSEQRNQTMVSPNSGDNNNQVMSSSPSSWYCHQTKTNGGLSVELLEENE
KLRSQNIQLNRELTMKSIDCNIYSLMSNYVGSQPTDRSYSPGSSSQPMEFLPAKRFSE

MEIEEEEEASPRFLGVPVIGLKRTRSEGVQVKTAVVGENSDEETPWLRYHNRNQRVCN*
>G1228 (63..1139)
GCATTTATAATTACTCACTCATCTTCTTTTCATTACATTACATACCAAACAAGAGCTCTC
AAATGGAAGGTTTCAAGGACACATCAACCCCTGTTCTTCGATCGAAAACCGGATGTGA
GAAGCCTCGAGGTTCAAGGATTTGCAGAGGCTCAAAGCTTTGCTTTCAAAGAAAAGAGG
AAGAAAGCTTACAAGATACAGTTCCATTTCTACAGATGCTGCAAAGTGAAGACCCCTCAT
CGTTTTTTTCAATCAAAGAGCCAACTTTCTGACGCTACTGTCTCTTCAAACCCCTCAAGG
AGCCTTGGGAAGCTCGAAAGATATCTTTCACTTGAGGATTACAAATTTCACTTACCAGTCC
AATCTGAGACCAACCGCTTCATGGAAGGAGCCAATCAAGCTGTGTCAAGCCAAGAAATTC
CCTTTAGCCAAGCAAACATGACACTCCCTTCTTCTACCTCATCACCCTCAGTGCACATT
CAAGACGAAAGCGCAAAATCAACCCTTGCTGCCTCAAGAAATGACTAGAGAAAAGAGAA
AGAGGAGGAAAACAAAACCAAGTAAAAACAATGAAGAGATTGAGAATCAAAGAATAAACC
ACATTGCTGTTGAACGAAACAGAAGAGCTCAAATGAACGAACATATCAACTCTCTCCGGG
CCCTTCTCCACCTTCTTACATCCAACGAGGAGACCAAGCTTCCATAGTAGGAGGAGCAA
TAACTACGTGAAGGCTCTCGAGCAAATCATACAATCTCTCGAATCGCAAAAGAGAACGC
AACAAACAAAGTAACAGTGAGGTAGTAGAAAACGCACTTAATCATCTCTCAGGCATTTCTG
CGAACGACCTGTGGCAACTCTTGAAGATCAAACCTGTATCCCCAAAATCGAAGCTACAG
TGATACAAAACCATGTGACGCTTAAAGTTCAATGTGAGAGAAACAAGGACAACCTTCTCA
AAGGAATCATATCACTTGAAAAGCTTAACTCACTGTTCTTCACTCTCAATATCACTACTT
CGTCTCATCTCTCTGTTTCTTATCTTCAACCTCAAGATGGAAGATGAGTGCAGCTTAG
AGTCAGCCGACGAGATTACGGCGCTGTTTCATCGGATTTTCGATATTCGACAATTTGAT
TAAACACATATAATTCCAAAATATTAACAGCTGACAAAATGGTATCTTTGCGGCC
>G1228 Amino Acid Sequence (domain in AA coordinates: 179-233)
MERFQGHINPCFFDRKPDVRSLEVQGFQFAEQSFAPKEKEEESLQDTPVPLQMLQSEDPSS
FFSIKEPNFLTLTSLQTLKEPWELERYLSLEDSQFHSVPQSETNRFMEGANQAVSSQEIP
FSQANMTLPSSSTSSPLSAHSRRKRKINHLPPQEMTREKRKRKTKPSKNNEEIEENQRINH
IAVERNRRRQMNHINSLRALLPYSYIQRGDQASIVGGAINYVKVLEQIIQSLESQKRTQ
QQNSEVVENALNHLGSISSNDLWTTLEDQTCIPKIEATVIQNHVSLKVQCEKKQGQLLK
GIIISLEKLKLTVLHLNITSSSHSSVSYSFNLKMEDECDLESADETAAVHRIFDIPTI*
>G1277 (51..512)
ATTCTAAAGTCCTCTCTCGGAAAGTAAGAGACTCAACTTCCGAGCCGCCATGGACGCCG
GAGTAGCAGTAAAAGCTGACGTGGCAGTCAAATGAAGAGAGAAAGACCATTCAAAGGGA
TCAGAATGAGAAAATGGGGGAATGGGTTGCGGAGATTCGAGAACCCAACAAGCGTTCAA
GACTTTGGCTCGGCTCTTACTCTACTCCGAAGCGCGCGCGTGCATACGACACGGCTG
TCTTTTACCTCAGAGGACCAACTGCTACGCTCAACTTCCCGAGCTTCTGCGGTGTACCT
CCGCCGAGGATATGTGACGCGCAACGATCAGGAAAAAGGCGACGGAGGTGGGAGCTCAAG
TAGATGCGATAGGGGCGACGGTGGTGCAGAACAAACGCGCGCGCTTTTGTAGTCAA
AGCGTGACTTTGGCGCGGGTTATTAGAGCTTGTGACTTGAACAAGTTACCTGACCCGG
AAAATCTCGATGATGATTGGTGGGAAAATAGACTGAAAAATAATAATAAATATCTTAC
AATGGTGGCTGTAGCTATCGTACGCGAATGCTTGGGCTTGTGTATATGACTACGTGGT
TACGGAAGATTCTCTGTTTCTGTCATTGTATTAAATTTAATCCACAAGTCAAACATA
CTGTACATTATTCTTAATTTAGTATTTTCTTATTAATATCTATCATTTGTTTGGTGAACA
CCAGAATATTAGACTATTAAATGTAACGAGTTTTTAATATTTTCGATCATAATAACACCAAG
CTAGTTAAAGGTTAATATCTTGTACGAGCTTGTAGTAAGTTCAATTGTATATATATG
TAACGGAAGAGGTTCTGTTCCGGTCCCAAGTGAAGTGATCAAAGGTGACTTCACATAAAA
AATAAAAAAAA
>G1277 Amino Acid Sequence (domain in AA coordinates: 18-85)
MDAGVAVKADVAVKMKRERPFKGIRMRKWKVABEIREPNKRSRLWLGSYSTPEAAARAY
DTAVFYLRGPTATLNFPELLPCTSAEDMSAATIRKKATEVGAQVDAIGATVVQNNKRRRV
FSQKRDFGGGLLELVBLNKLDPENLDDDLVGK*
>G1309 (53..859)
CGTCGACCTCTTAATTAAGACGACTTGAGAGAGAAAGAAAGATACGTGGAAGATGACCAA
ATCTGGAGAGAGACCAAAAACAGAGACAGAGAAAGGGTTATGGTCACCTGAAGAAGACCA
GAAGCTCAAGAGTTTCATCTCTCTCGTGGCCATGCTTGTGAGGACCTGTTCCCATCT
AGCTGGATTGCAAAGGAATGGGAAAAGCTGCAGATTAAGGTGGATTAAATTACCTAAGACC
AGGACTAAAGAGGGGGTCTGTTTAGTGAAGAAGAAGAAGAGACCATCTTGACTTTACATTC
TTCCTTGGGTAACAAGTGGTCTCGGATTGCAAAATATTTACCGGAAGAACAGACAACGA

GATTAAGAACTATTGGCATTCTCTATCTGAAGAAGAGATGGCTCAAATCTCAACCACAACCT
CAAAAGCCAAATATCAGACCTCACAGAATCTCCTTCTTCACTACTTCTTGCGGGAAAAG
AAATCTGGAAACCGAAACCTTAGATCACGTGATCTCCTTCCAGAAATTTTCAGAGAATCC
AACTTCATCACCATCCAAAGAAAGCAACAACAACATGATCATGAACAACAGTAATAACTT
GCCTAAACTGTTCTTCTCTGAGTGGATCAGTTCTTCAAATCCACACATCGATTACTCCTC
TGCTTTTACAGATTCCAAGCACATTAATGAAACTCAAGATCAAATCAATGAAGAGGAAGT
GATGATGATCAATAACAACAACACTACTCTTCACTTGAGGATGTCATGCTCCGTACAGATTT
TTGTCAGCCTGATCATGAATATGCAAATATTATTCTTCTGAGATTTCTTCATCAACAG
TGACCAAAATTATGCTCTAAGAAGAGTGAATATGATCGTAAGAGGAACATAAGCTAGTTAC
TTGTGTTACAGC

>G1309 Amino Acid Sequence (domain in AA coordinates: 9-114)
MTKSGERPQKQKGLWSPEEDQKLKSFILSRGHACWTTVPILAGLQRNGKSCRLRWINY
LRPGLKRGSPFSEEEETILTLHSSLGNKWSRIAKYLPGRDNEIKNYWHSYLKKRWLKSQ
PQLKSQISDLTESPSSLLSCGKRNLLETLDHVISFQKFSENPTSSPSKESNNNMIMNNS
NNLPKLFSEWISSSNPHIDYSSAFTDSKHINETQDQINEEVMNNNNYSSLEDVMLR
TDFLQPDHEYANYSSGDFFINSDQNYV*

>G1314 (1..990)
ATGGGAAGAGCTCCGTGTTGCGACAAGACAAAAGTGAAGCGAGGGCCTTGCTCGCCTGAA
GAAGACTCTAAACTTAGAGATTACATTGAAAAGTATGGTAATGGTGGAATTTGGATCTCT
TTCCCCCTCAAAGCCGGTTTGAGGAGATGTGGGAAGAGTTGTAGACTGAGGTGGCTAAAC
TATTTGAGACCAACATAAAGCATGGTGACTTCTCTGAGGAAGAAGACAGGATCATTTTT
AGTCTCTTCGCTGCCATAGGAAGCAGGTGGTCAATAATAGCAGCTCATCTACCGGGACGA
ACAGACAACGACATAAAAACTATTGGAACACAAAGCTAAGGAAGAACTCTTGCTCTCT
TCCTCTGATTCATCATCATCAGCCATGGCTTCTCTTATCTAAACCCTATTTCTCAGGAT
GTGAAAAGACCAACCTCACCAACAACAATCCCATCTTCTTCTTACAATCCGTATGCTGAA
AACCTTAATCAATACCCAACAATAATCCCTCATCTCCAGCATCAATGGCTTCGAAGCTGGT
GACAAAACAGATAATTTCTATATTAACCCTAATTATCCTCAAGATCTCTATCTCTCGGAC
AGCAACAACAACCTCGAAGCAGCAATGGTTTCTTGCTCAACCACAATATGTGTGATCAG
TACAAGAACCACACAGTTTTTCTTCAGACGTCATGGGATAAGATCAGAGATTATGATG
AAGCAAGAAGAGATAATGATGATGATGATGATAGACCACCACATTGACCAGAGGACAAAA
GGGTACAATGGGGAATTCACACAAGGGTATTATAATTACTACAATGGGCATGGGGATTG
AAGCAAATGATTAGTGGGAACAGGCACTAATTCTAACATAAACATGGGTGGTTCAAGTTCA
TCTTCTAGTTTCGATAAGCAACCTAGCTGAGAACAAAAGCAGTGGTAGCCTCCTACTAGAA
TACAAATGCTTGCCCTATTTCTACTCCTAG

>G1314 Amino Acid Sequence (domain in AA coordinates: 14-116)
MGRAPCCDKTKVKGFPWSPEEDSKLRDYIEKYNGNGNWFPLKAGLRRCGKSCRLRLN
YLRPNIKHGFSEEDRIIFSLFAAIGSRWSIIAHLPGRTDNDIKNYWNTKLKRLKLLSS
SSDSSSSAMASPYLNPISQDVKRPTSPPTIPSSSYNPAENPNQYPTKSLISSINGFEAG
DKQIIISYINPNYPQDLYLSDSNNTSNANGFLLNHNMDQYKNHTSFSSDVNGIRSEIMM
KQEEIMMMMHIDQRTKGYNGEFTQGYNYNNGHDLKQISMISGTGTNSNINMGSGS
SSSSISNLAENKSSGSLLEKCLPYFYS*

>G1317 (1..849)
ATGGGAAGATCACCTGTTGTGATAAAAAATGGAGTGAAGAAGGGACCATGGACTGCTGAG
GAGGATCAGAAACTCATCGATTATATTCGATTTCATGGTCCTGGCAATTGGCGTACGCTC
CCCAAAAATGCTGGACTCCATAGATGTGGAAAAGCTGCCGTCTTCGATGGACCAATTAT
CTAAGACCGGACATCAAGAGAGGAAGATTCTCGTTTCGAGGAAGAAGAACTATCATTAG
CTACACAGTGTATGGGAAACAAGTGGTCAGCAATAGCCGCTCGTCTACCAGGGAGGACC
GATAACGAAATAAAAAACCATTTGGAACACTCACATCCGCAAGAGACTTGTAAGGAGTGGT
ATCGACCTGTTACTCATTTCTCCACGCCTTGATCTTCTTGATTGTCTCACTTTTGAGT
GCACTTTTCAACCAGCCAACTTTTCAGCAGTTGCAACACATGCGTCTTCTCTCTTAAT
CCTGATGTATTGAGGTTGGCCTCTCTACTACTGCCACTTCAAAACCTAATCCAGTTTAC
CCATCGAACCTCGACCAAAATCTTCAAACTCCAAATACATCATCAGAATCGTCTCAACCA
CAAGCTGAGACTAGTACAGTCCCAACAACTATGAAACTTTCATCATTGGAGCCTATGAAC
GCAAGACTCGACGAGCTTGGTCTTGACAGATGTATTACCACCTTTGTCAGAGAGTTTGTAC
TTAGACTCGCTCATGTCAACGCCAATGTCTTCTCCACGACAAAATAGCATTGAAGCAGAA
ACCAACTCCAGCACTTTCTTCGACTTTGGAATTCCGGAAGATTTTCATCTTAGATGACTTT
ATGTTTTAA

>G1317 Amino Acid Sequence (conserved domain in AA coordinates:13-118)
MGRSPCCDKNGVKGPWTAEDQKLIDYIRFHGPGNWRTLPLKNAGLHRCGKSCRLRWITNY
LRPDIKGRGFSFEEETIIQLHSVMGNKWSAIAARLPGRTDNEIKNHNTHIRKRLVRS
IDPVTHSPRLDLLLSSLLSALFNQPNFSAVATHASSLLNPDVLRASLLPLQNPVPVY
PSNLDQNLQTPNTSSSQPQAETSTVPTNYETSSLEPMNARLDDVGLADVLPLPSESFD
LDSLMSTPMSSPRQNSIEAETNSSTFFDFGIPEDFILDDFMF*

>G1323 (49..870)

AAGAGGGAATCTCAAAAGTGTGTCTGTGTGAGAGAGGAGAGAGAATATGGGCAAAGGA
AGAGCACCATGTTGTGACAAAACCAAGTGAAGAGAGGACCATGGAGCCATGATGAAGAC
TTGAAACTCATCTCTTTTCATTCAAGAATGGTCATGAGAATTGGAGATCTCTCCCAAAG
CAAGCTGGATTGTTGAGGTGTGGCAAGAGTTGTCGTCTGCGATGGATTAATTACCTCAGA
CCTGATGTGAAACGTGGCAATTTCAAGTGCAGAGGAAGAAGACACCATCATCAAACTTCAC
CAGAGCTTTGGTAAACAAGTGGTCAAGATTGCTTCTAAGCTGCCTGGAAGAAGACAGACAT
GAGATCAAGAATGTGTGGCATAACATCTCAAGAAAAGATTGAGCTCGGAAACTAACCTT
AATGCCGATGAAGCGGGTTCAAAGGTTCTTTGAATGAAGAAGAGAAGTCTCAAGAGTCA
TCTCCAAATGCTTCAATGTCTTTTGTCTGGTTCCAACATTTCAAGCAAAGACGATGATGCA
CAGATAAGTCAAATGTTTGAAGCATTCTTAAGTATAGCGAGTTTACGGGGATGTTACAA
GAGGTAGACAAACCAGAGCTGCTGGAGATGCCTTTTGAATTAGATCTGACATTTGGAGT
TTCATAGATGGTTCAAGTCAATCTCAACAACAGAGAACAGAGCTCTTCAAGAGTCTGAA
GAAGATGAAGTTGATAAATGGTTTAAGCACCTGGAAGCGAACTCGGGTTAGAAGAAAAC
GATAACCAACAACAACAGCATAAACAGGGAACAGAAAGTGAACATTCATCATCACTC
TTGGAGAGTTACGAGCTCCTCATACATTAATGAAGCCATAAAGCAAGTCAATTTTACCTT
GAAAATGGAATTATTAGCTAAGTATTGTCATTATTAGTATATAAGCAAGATCAGATAGG
CGCATGTAGTAGCAACAAGAAACGTCGAATTGTAGACAAAATGTAGATATTACAGA
GTTGAAAGATTGTATTTTGCAAATGATTGCTTTGTAGTGAAATCAAGTTATCACAAAAA
AAAAAAA

>G1323 Amino Acid Sequence (domain in AA coordinates: 15-116)
MGKGRAPCCDKTKVKRGPWSHDEDLKLISFIHKNHENWRSPLKQAGLLRCGKSCRLRWI
NYLRPDVKRGNFSAEEETIIKLHQSFGNKWSKIASKLPGRTDNEIKNVWHTHLKKRLSS
ETNLNADEAGSKGSLNEEENSQESSPNASMSFAGSNISSKDDDAQISQMFHILTYSEFT
GMLQEVDPKPELLEMPFDLDPDIWSFIDGSDSFQQPENRALQSESEDEVDKWFHLESELG
LEENDNQQQQHKQGTEDHSSSLLESYELLIH*

>G1332 (1..606)

ATGGAATGCAAAAGAGAAGAAGGGAAGTCTTACGTGAAGAGAGGGTTGTGGAACAGAA
GAAGATATGATATTAAAAAGCTATGTTGAGACTCATGGTGAAGGAAACTGGGCAGACATT
TCTCGTAGATCCGGGTGAAGAGAGGAGGAAAAAGCTGTAGGCTGAGATGGAAGAAGTAT
CTAAGACCAAATATCAAAAGAGGAAGCATGTCAACCAAGAACAAGACCTTATCATCCGC
ATGCATAAGCTTCTTGAAACAGATGGTTCGTTGATCGCTGGCTCCAGGTCGTACT
GACAATGAAGTGAAGAACTACTGGAATACTCATTTGAACAAGAAACCTAATCCCCGAAA
CAGAATGCACCTGAATCAATCGTCGGCGCCACTCCTTTCAAGGATAAGCCAGTTATGTCT
ACAGAACTGAGAAGAAGCCATGGAGAAGGAGGAGAAGAGGAGCAATACCTGGATGGAG
GAGACCAACCACTTTGGCTATGACGTCCACGTAGGATCTCCCTTGCCACTTATTTCCAC
TACCCAGACAACACTCTCGTGTTTGACCCATGTTTTTCCTTTACCGATTCTTTCTCTCTG
CTTTAG

>G1332 Amino Acid Sequence (conserved domain in AA coordinates:13-116)
MECKREEGKSYVKRGLWKPEEDMILKSYVETHGEGNWADISRRSGLKRGKSCRLRWKNY
LRPNIKRGSMSPPQEQDLIIRMHKLGNRWSLIAGRLPGRTDNEVKNYWNTHLNKKPNRSK
QNAPEIVGATPFTBKPVMSTELRRSHGEGGEEESNTWMEETNHFGYDVHVGSPPLISH
YPDNTLVFDPFCSFTDFPPL*

>G1334 (76..885)

ATAGCTCCCAACTAATAGGAATCTCAAGCTTCTCACTCTCTCTGTTTTTCCATTGGACT
TTTGGAAACATAAGCTATGCAAACTGAGGAGCTTTTGTGCGCCACCACAGACTCCTTGGTGG
AATGCTTTTGGATCTCAGCCGTTGACTACAGAGAGCCTTTCCGGCGAAGCTTCTGATTCA
TTCACCGGAGTTAAGGCAGTTACTACGGAGGCAGAACAAAGGTGTGGTGGATAAACAACT
TCTACAACCTCTCTTCACTTCTCACCTGGTGGTGAAGAGATTCAAGAGATGTGCCAAAG
CCTCATGTTGCTTTGCGGATGCAATCAGCTTGCTTCGAGTTTGGATTGTCTCAGCCAATG
ATGTACACAAAGCATCCTCATGTTGAACAATACTATGGAGTTGTTTCAGCATACGGATCT

CAGAGGTCTTCGGGCGGAGTAATGATTCCACTGAAGATGGAGACAGAAGAAGATGGTACC
ATCTATGTGAACCTCAAAGCAGTACCATGGAATTATCAGGCGACGCCAGTCCCGAGCAAAG
GCTGAAAAAAGTGAAGTAGATGCCGTAAGCCATATATGCATCACTCAGGCCATCTCCATGCT
ATGCGCCGCTCTAGAGGATCTGGCGGGCGTTTCTTGAACACCAAGACAGCTGATGCGGCT
AAGCAGTCTAAGCCGAGTAATTCTCAGAGTCTGAAGTCTTTCATCCGAAAAATGAGACC
ATAAACTCATCGAGGGAAGCAAATGAGTCAAATCTCTCGGATTCTGCAGTTACAAGTATG
GATTACTTCTAAGTTCGTGGGCTTATTCTCCTGGTGGCATGGTCATGCCTATCAAGTGG
AATGCAGCAGCAATGGATATTGGCTGCTGCAAACTTAATATATGATCAGCAGATAGGGGA
CAAGACATGATTGGTCACCAAGTCCCTTTGTCTTGTCCCTTATCTTTCAGCCAAACGGAAA
GAGAACTTGTCTTGGAAAAAAGACATTGAGTTTCCTTGGTTTATAAGATTGGTCCTTT
TACCATCCCTTTGGCTGTAAACAGGCAAATCATCTTTGGCTCATGCTTCATCAAGTCTT
ATCTTCGTCTGTTTTCTTCTACGCATCTTCATAAGATCTCTGAAGTAGTGAATAACATTT
CCTAGCATCATGTTTCAACTAGTGTGTGTGTGAAGAACTCTGCCTATTTCCAGATGAT
GTATTGTGTGTAACGTGTTTATGAAACAAACGTAAGACTTTCAAGTTAAAAA
AAAAAAAAAAAAA

>G1334 Amino Acid Sequence (domain in AA coordinates: 18-190)

MQTBELLSPQTPWNAFGSQPLTTESLSGEASDSFTGVKAVTTEAEQGVVDKQTSTTLF
TFSPGGEKSSRDVDPKPHVAFAMQSACFEFGFAQPMYTKHPHVEQYGVVSAYGSQRSSG
RVMIPKMETEEDGTIYVNSQYHGIIRRRQSRKAELSRCKPYMHHSRHLHAMRRPR
GSGGRFLNKTADAQSKPSNSQSSEVFHPENETINSSREANESNLSDSAVTSMDYFLS
SSAYSPGGMVMPKWNAAAMDIGCCKLNI*

>G1381 (32..802)

CAGCTTTAACTACTCTCTCTCTCTCAATGGGAAAAAACAATCAACATAGAGAGTAG
TGCTACTCATCATCAAGACAATATTGTTTCCGTTATAACAGCCACGATATCCTCCTCCTC
CGTCGTAACGCTCTTCGTAGACTCTTGGTCTACCTCCAAAAGATCGTTAGTGCAAGACAA
TGACTCCGGAGGGAAACGGCGGAAGAGCAACGTTAGTGATGATAACAAGAAATCCGACGTC
GTATAGAGGAGTGAGGATGAGGAGTTGGGGAAAATGGGTGTCGGAGATTAGAGAGCCGAG
GAAGAAATCAAGAAATATGGCTTTGGCACTTATCCAACGGCAGAGATGGCAGCTCGTGCTCA
TGATGTGGCGGCTTTAGCTATTAAAGGCAACTCCGGTTTTCTTAATTTCCCTGAATTATC
CGGTTTGCTTCCCTCGTCCGTTAGCTGCTCTCTTAAGGATATAACAAGCTGCAGCTACCAA
AGCCGCGAAGCAACCAACGTCGACAAACCGGTTATCGATAAGAAATTAGCTGATGAGCT
AAGCCACTCTGAGTTGTTGTCTACCGCTCAGTCTTCGACTTCTAGTAGTTTCGTGTTTTCT
TTCGGACACGTCGGAGACTTCTAGTACGGACAAGGAAAGCAACGAAGAGACGGTGTGTTGA
TTTGCCGGACCTTTTACGGACGGGCTTATGAACCCAAACGATGCGTTTTGTTTATGCAA
CGGCACCTTTACGTGGCAGCTTTACGGAGAGGAGGATGTAGGGTTCAGGTTTGAAGAGCC
GTTTAATTGGCAAAATGACTAAACCGCCCTCCACTTGCTTACTGTAATTACTAACATATA
ATTTTCTTGATAAAGAACATATATTTCCATTACGGTATTAACTAATCTTTTCTATCCTTT
TCTCTTTTCTTGTTTCTACATCTGAGTATATTGTCACTATGTGAAAAAATTGATCTCGTT
TTGAATATTTACTTTTCAAAATTGAAGTAACGCAAGTGATTGATAAAAAAAAAAAAAA

>G1381 Amino Acid Sequence (domain in AA coordinates: TBD)

MGKQINIESSATHHQDNIVSVITATISSSSVVTSSSDSWSTSKRSLVQDNDSGGKRRKSN
VSDDNKNPTSIRGVRMRWSGKWVSEIREPRKKSRIWLGYPTAEMAARAHDAALAIKGN
SGFLNFPPELGLLRPVSCSPKDIQAAATKAAEATTWHKPVIDKKLADELHSELLSTAQ
SSTSSSFVFSSTSETSTDKESNEETVFDLPDLFTDGLMNPNDADFCLCNGTFTWQLYGE
EDVGFRFEPPFNWQND*

>G1382 (90..1763)

CTCTCATTTCGCCATAGCTGAGAGCTTCTTCTACTTTCCCTTAGCTTCTTTTTTCCTTCA
TTTTTGTCTACCCFTGCGAATCTCTGAAATGAACCTCAAGCTAATGACCGGAAGGAGT
TTCAGGGAGATTGTTTCGGCGACGGGAGATCTCACGGCAAAGCACGATTGAGCTGGAGGAA
ACGGAGGTGGAGGTGCTAGGTATAAGCTGATGTACCGGCCAAGCTTCCGATCTCGAGGT
CGACTGATATCAGATTCTCTGGGTTGAGTCCGACTTCGTTTTTGGAAATCTCCTGTTT
TCATCTCAACATCAAGCCAGAACCCTTCCCTACTACTGGTCTTTGTTCAAGCCTCGAC
CAGTGCAATTTCTGTAGCTCAAGTTCTTATACAGGCAGGGGGTTCCATCAGAACACCT
TTACTGAGCAGAAGTCCAGTGAATTTGAGTTCAGACCTCCTGCATCAAATATGGTATATG
CAGAGCTTGGCAAGATTAGAAGTGAGCCACCAGTACATTTTCAAGGCCAGGGCCATGGAT
CCTCACACTACCTTCTTCGATCAGTGATGCTGCAGGTTCTCAAGTGAGCTAAGCCGGC
CAACTCCTCCTTGTCAGATGACACCAACGAGCTCAGATATTCGGCTGGATCTGATCAAG

AGGAATCAATCCAGACTTCCCAAAATGACTCCAGAGGAAGCACTCCATCCATCTTGGCTG
ATGATGGTTATAACTGGAGAAAATATGGTCAAAAGCATGTCAAAGGGAGTGAATTTCCCC
GGAGCTATTATAAATGTACACATCCTAATTGTGAAGTGAAAAAGTTATTTGAAAGATCTC
ATGATGGGCAGATCACCAGATATTATATACAAGGGTACACATGACCATCCTAAACCTCAAC
CTGGTCGCCGAACTCTGGTGGTATGGCTGCACAAGAAGAAAGGCTAGACAAGTATCCTT
CTTCAACTGGCCGAGATGAGAAGGGATCTGGCGTCTACAACCTGTCTAACCCCAATGAAC
AACTGGTAACCTGAAGTACCTCCTATCTCAGCATCTGACGATGGTGGAGAAGCGGCAG
CGTCAAATAGGAATAAAGATGAGCCGGACGATGATGATCCATTCTCAAAACGGAGGAGGA
TGGAGGGTGGCATGGAATAAATCCACTAGTGAAACCCATCCGGGAGCCTCGGGTTGTTG
TTCAAACCTCTGAGTGAAGTTGACATCTCTGGATGATGGTTATAGATGGCGCAAAATATGGGC
AGAAAGTCGTAAGGGGGAACCCAAATCCCAGGAGCTACTACAAATGCACAGCTCATGGAT
GCCAGTGAGAAAACACGTGGAGAGAGCATCACATGATCCAAAGCTGTAATAACAACAT
ACGAAGGCAAAACACGATCATGATGTTCCCACTTCAAAGTCTAGCAGCAATCACGAAATCC
AGCCTCGGTTAGACACGATGAAACAGACACCATCAGCCTCAATCTTGGTGTGGAATCT
CATCTGATGGACCTAACACGCTTCCAACGAACATCAGCACCAGAATCAACAACCTGTCA
ACCAAACCTCACCCAAATGGAGTCAATTTTCAAGTTTGTTCATGCTAGTCCCATGTCATCCT
ACTATGCTAGCTTAAATAGCGGTATGAATCAGTACGGCCAGAGAGAAACAAAGAACGAGA
CTCAAAATGGTGACATCTCGTCTTGAACAATTCATCTTACCCATATCCGCCCAACATGG
GGAGGTACAATCGGGTCCGTAAAAACAAAAGTAAGCAACATTATGTACGGGATCTTCTT
AGGTTAGGAATGGGACGAGGCTTGTCTATATAATTCCTATTTCTTACAGAGAGCTGA
TCTTGATTCAAATATCTCCACCATATATATTTGTTTGTGTCACCTGTATTGAGTTCCAA
AAATGTTATGTAAAAATACACAACAAGATGTTAATGCTTTTATTTAAACAAGAAACAGCA
ATATTACTACAAAAAATAAAAAAAAAA
>G1382 Amino Acid Sequence (domain in AA coordinates: 210-266, 385-437)
MNPQANDRKEFQGD SATGDLTA KHDSAGNGGGGARYKLMSPAKLPISRSTDI TTPGL
SPTSFLSPVFI SNIKPEPSPTT GSLFKPRPVHISASSSSYTGRGFHQNTFTEQKSSEFE
FRPPASNVMVYAE LKIRSEPPVHFQGGHSSHPSSISDAAGSSSEL SRPTPPCQMTPT
SSDIPAGSDQEE SIQTSQNSRGS TPSILADDGYNWRKYGQKHVKGSEFPRSYKCTHPN
CEVKKLFERSHDGQITDI IYKGT HDHPKQPGRNRNSGMAAQEERLDKYPSS TGRDEKGS
GVYNLSNPNEQTGNPEV PPI SASDDGGEAAASNRNKDEPDDDDPFSKRRRMEGAMEITPL
VKPIREPRVVVQTLSEVD ILDDGYRWRKYGQKVVRGNPNPRSYKCTAHGCPVRKHVERA
SHDPKAVITTYEGKHDH DVP TSKSSSNHEIQPRFRPDETD TISLNLGVGISSDGP NHASN
EHQHQNQLV NQTHPNGVNFRFVHASPMSSYYASLNSGMNQYQRET KNETQNGDISSLN
NSSYPYPPNMGRVQSGP*
>G1435 (8..904)
GTGAAACATGGGGAAGGAAGTTATGGTGAGCGATTACGGTGACGACGAGAGAGACGC
CGGCGGCGCGATGAATATAGGATTCCGGAATGGGAAATTGGTTTACCCAACGGAGATGA
TTTGACTCCGTTATCTCAATATCTAGTCCCGTCGATTCTCGCGTTAGCTTTCAGCATGAT
CCGAGAACGAAGCCGTACAATTCACGACGTCAATCGCGCGTCGCAATCACGCTCTCTTC
GTTGAGAAGCAGTACCAATGCTTCGTCTGTGATGGAGGAGGTCGTGGATCGAGTTGAATC
GAGTGTTCAGGATCAGATCCGAAGAAACAGAAGAAATCGGATGGTGGTGAAGCAGCGGC
GGTGGAGGATTCCACGGCGGAGGAAGGAGACTCCGGGCCTGAAGACGCGTCTGGGAAGAC
ATCGAAACGACCGGTTTAGTGTGGACACCGCAGCTACACAAGAGATTGTGGACGTTGT
GGCTCATCTAGGGATTAAAAACGCA GTGCCGAAGACGATTATGCAGCTGATGAACGTGGA
AGGACTTACTCGTGAGAACGTTGCGTCTCATTTGCAGAAATATAGGCTTTACCTTAAACG
GATTCAAGGATTGACGACGGAAGAAGATCCTTATTCTGTCGTCGGATCAGCTCTTCTCTTC
AACGCCGGTTCTCTCCACAGAGCTTTCAAGACGGCGGAGGAAGTAACGGAAGTTGGGGGT
TCCGGTTCCGGTTCEGTGATGGTGCTTATCCAGGCTATGGGAATCAAATGGGTATGCA
AGGATATTATCAACAGTATAGTAACCATGGCAATGAATCAAACCAATATATGATGCAGCA
GAATAAGTTTGAACAATGGTGACATATCCTTCTGTTGGTGGTGGTGACGTGAATGACAA
GTAAATGGATCTTAAAGGTCTATAATTTGCTCTACAGAGAGATACTGGTTCTTGGCTTAT
GGTTTATTTTCCCACTTCATGAGGTTGTTGTGACTTTTAATTCTCCATGTTTTCCACACA
AGTCTTTATTGCTTTTGTATAGAAAATGATTTGAGAAAATCACTGGGAAGCTTGGTATT
GTTGGAGGATGAAGCCTTCTATGAATGATTTAGTTTCTACTGTCTCCATTCTTTATGAG
GTAATAAAGCCTTCTTTTGCTCATCGCTTGTAGTCTTCTTAAATTCAAGACAGCGTCACA
TGTTTGTTCGGTTATGTTAATGTTTCTTTCTTTGGATAATGAAGATAGCATCAGGTCTC
ATGTCTCCTCACTTTGATAAA

>G1435 Amino Acid Sequence (domain in AA coordinates: 146-194)
MGKEVMVSDYGDGDDGDEYRIPEWEIGLPLNGDDLTPLSQYLVP SILALAFSMIPE
RSRTIHDVNRASQITLSSLRSSTNASSVMEEVVDREVSVPGSDPKKQKSDGGEAAAVE
DSTAEEGDSGPEDASGKTSKRPLRVWTPQLHKRFVDVVAHLGIKNAVPKTIMQLMNVEGL
TRENVAHLQKYRLYLKRIQGLTTEEDPYSSSDQLFSSTPVPPQSPQDGGGSGKLGVPV
PVPSMVPIPGYGNQMGMQGYQQYQYNSHGNESNQYMMQKNKFGTMVTYPSVGGGDVNDK*

>G1537 (1..783)

ATGGAACCGAAGTAAACGAGGAACAGCAAGCAGTTCAAGATGGAACCCAACGAAAGAT
CAGATCACGCTACTGGAAAATCTTTACAAGGAAGGAATACGAACCTCCGAGCGCCGATCAG
ATTACAGCAGATCACCGGTAGGCTTCGTGCGTACGGCCATATCGAAGGTAAAAACGCTCTT
TACTGGTTCAGAACCATTAAGGCTAGGCAACGCCAAAAGCAGAAACAGGAGCGCATGGCT
TACTTCAATCGCCTCCTCCACAAAACCTCCCGTTTCTTCTACCCCCCTCCTTGCTCAAAC
GTGGGTTGTGTGCTAGTCCGTACTATTTACAGCAAGCAAGTGATCATATGAATCAACAT
GGAAGTGTATACACAAACGATCTTCTTACAGAAACAATGTGATGATTCCAAGTGGTGGC
TACGAGAAACGGACAGTCAACAACATCAGAAACAACCTTTCAGACATAAGAACAACAGCA
GCCACAAGAAATGCCAATTCTCCGAGTTCAGTCAAGTTTGACAGATTGCCCCCTCCGTGAT
AACTGTTATGCCGGTGAGGACATTAACGCTCAATTCCAGTGGACGGAACACTCCCTCTT
TTTCTCTTTCAGCCTTTTGAATGCAAGTAATGCTGATGGTATGGGAAGTTCCAGTTTGGC
CTTGGTAGTGATTCTCCGGTGGATTGTTCTAGCGATGGAGCCGGCCGAGAGCAGCCGTTT
ATTGATTCTCTTTCTGGTGGTCTACTTCTACTCGTTTCGATAGTAATGGTAATGGGTTG
TAA

>G1537 Amino Acid Sequence (domain in AA coordinates: 14-74)
MENEVNAGTASSSRWNPTKDQITLLENLYKEGIRTPSADQIQITGRLRAYGHIEGKNVF
YWFQNHKARQRQKQKQERMAFNRLLHKTSRFFYPPPCSNVGCVSPYYLQOASDHMMNQH
GSVYTNDDLHRNNMIPSGGYEKRTVTQHQKQLSDIRTTAATRMPISPSLRFDRLALRD
NCYAGEDINVNSSGRKTLPLFPLQPLNASNADGMGSSSFALGSDSPVDCSSDGAGREQPF
IDFFSGGSTSTRFDSNNGNL*

>G1545 (67..729)

CATCACCAATCTTTGAATCTAAGAGAGAGAAGAAGAAGGTCTAGAGAACGAAAAGA
AGAAACATGAATAACCGAATGTAGATGATCATAATCTTCTACTCATTTCTCAATTGTAC
CCTAATGTCTATACTCCATTAGTACCACAACAAGGAGGAGAAGCAAAACCAACACGGCGG
AGGAAAAGGAAGAGCAAGAGTGTGTGGTGGCAGAGGAGGGTGAAAACGAAGGCAATGGG
TGGTTTAGAAAGAGAAAATTGAGTGATGAGCAAGTAAGAATGTTGGAGATTAGCTTTGAA
GACGATCATAAGCTTGAATCCGAGAGGAAAGATCGGCTTGCTTCTGAGTTAGGGCTTGAT
CCTCGTCAAGTCGCGCTCTGGTTCCAAAACCGCGTGACGGTGGAAGAACAACGAGTC
GAGGATGAATACACTAAACTCAAGATGCATACGAAACCACCGTCGTTGAGAAATGTCTGT
CTTGATTCTGAGGTTATTACCTAAAGGAACAACCTTTACGAGGCTGAAAGAGAGATCCAA
CGGCTTGCAAAAAGAGTTGAAGAACTTTAAGTAACAGTCCTATCTCATCTCTGTGACC
ATTGAAGCCAATCATACGACACCGTTTTTTTGGAGATTACGACATCGGATTTGACGGTGAG
GCTGACGAGAACTTGCTCTACTCGCCAGATTACATTGATGGATTAGACTGGATGAGCCAA
TTTATGTAAAAAACTATAAGCTAATCTATTTTTCAGTCGTAGTATAG

>G1545 Amino Acid Sequence (domain in AA coordinates: 54-117)
MNNQNVDDHNNLLISQLYPNVYTPLVPQQGGEAKPTRRRKRKSKSVVVAEEGENEGNGWF
RKRKLSDEQVRMLEISFEDDHKLESERKDRLASLGLDPRQVAVWFQNRARWKNKRVED
EYTKLKNAYETTVEKCRLDSEVIHLKEQLYEAREIQRLAKRVEGTLNSPISSSVTIE
ANHHTPPFGDYDIGFDGEADENLLYSPDYIDGLDWSQM*

>G1641 (1..867)

ATGGAGGTTATGAGACCGTCGACGTACACGTGTCAGGTGGGAACCTGGCTCATGGAGGAA
ACTAAGAGCGGCGTCGAGCTTCTGGTGAAGGTGCCACGTGGACGGCGGCAGAGAACAAG
GCATTCGAGAATGCTTTGGCGGTTTACGACGACAACACTCCTGATCGGTGGCAGAAGGTG
GCTGCGGTGATTCCGGGGAAGACAGTGAGTGACGTAATTAGACAGTATAACGATTGGAA
GCTGATGTCAGCAGCATCGAGGCCGTTTAAATCCCGGTCCCCGGTTACATCACCTCGCCG
CCTTTCACTCTAGATTGGCCCGGCGGCTGGCGGATGTAACGGGTTTAAACCGGTCAT
CAGGTTTGTAAATAAACGGTCGAGGCCGGTAGATCGCCGGAGCTGGAGCGGAAGAAAGGC
GTTCTTGGACGGAGGAAGAACAAGCTATTTCTAATGGGTTTGAAGAAATATGGGAAA
GGAGATTGGAGAAACATATCTCGGAACCTTGTGATAACGCGAACGCAACACAAGTAGCT
AGCCACGCCCAAAAGTACTTCATCCGGCAACTTTCCGGCGGCAAGGACAAGAGACGAGCA

AGCATTACGACATAACCACCGTAAATCTCGAAGAGGAGGCTTCTTTGGAGACCAATAAG
AGCTCCATTGTTGTTGGAGATCAGCGTTCAAGGCTAACCGCGTTTCCTTGGAAACCAAACG
GACAACAATGGAACACAGGCAGACGCTTCAATATAACGATTGGAAACGCTATTAGTGGC
GTTCAATTCATACGGCCAGGTTATGATTGGAGGGTATAACAATGCAGATTCTTGCTATGAC
GCCCCAAACACAATGTTTCAACTATAG
>G1641 Amino Acid Sequence (domain in AA coordinates: 139-200)
MEVMRPSTSHVSGGNWLMEEETKSGVAASGEGATWTAENKAFENALAVYDDNTPDRWQKV
AAVIPGKTVSDVIRQYNDLEADVSSIEAGLIPVPGYITSPFFTLWDAGGGGGCNGFKPGH
QVCNKRSQLGRSPELERKKGVPTWEEHKLFLMGLKKYKGDWRNISRNFVITRPTQVA
SHAQKYFIRQLSGGKDKRRASIHDIITVNLEEEASLETNKSSIVVDQRSRLTAFPNWQT
DNNGTQADAFNITIGNAISGVHSYGQVMIGGYNNADSCYDAQNTMFQL*

>G165 (19..699)
CTTCAAAACATCTAAAAAATGGTGAAAAAACTCTTGGTCGTAGAAAGGTAGAGATAGTG
AAAATGACTAAGGAATCAAAACCTTCAAGTCACATTTTCCAAGAGAAAAGCTGGTCTTTTT
AAGAAGGCTAGTGAATTTTGACATTTATGTGATGCAAAAATTCGATGATCGTGTTTTCA
CCAGCTGGAAGTATTTTCTTTTGGTCATCCAAATGTTGATGTTCTGCTTGACCACTTT
CGAGGGTGTGTTGTAGGACACAACAACACAACCTTGATGAAAGCTACACAAAGCTTCAT
GTTCAAAATGCTCAACAAATCCTACACTGAGGTGAAGGCGGAAGTAGAAAAAGAACAAAAG
AATAAGCAGTCGCGGGCTCAAAATGAAAGAGAAAACGAAAACGCTGAGGAGTGGTGGAGT
AAGTCTCCATTAGAACTCAACTTAAGTCAATCAACCTGTATGATACGTGTTCTTAAAGAT
TTGAAGAAGATAGTTGATGAAAAAGCAATTCAATTAATCCATCAAACAAACCCAACTTC
TATGTTGGAAGTCTAGCAATGCTGCTGCTCCAGCACTGTTAGTGGTGGTAATATCTCC
ACAAACCAGGGGTTCTTTGATCAAAACGGAATGACGACTAATCCTACTCAAACACTTCTG
TTTGGATTTGATATTATGAATCGCACACCAGGAGTTTAAATAAGTCTATCCTCATTATGG
GTCTTGGTACTATAAGTTCATCTCTCTCGTTGTTGACTTTTAAAGTCTCCAATAGTTTGT
TGTTG

>G165 Amino Acid Sequence (conserved domain in AA coordinates: 7-62)
MVKKTLGRRKVEIVKMTKESNLQVTFSKRKAGLFKKASEFCTLCDAKIAMIVFSPAGKVF
SFGHPNVDVLLDHFRCVVGHNNTNLDES YTKLHVQMLNKS YTEVKA EVEKEQKNKQSRA
QNERENENAE EWWSKSPLELNLQSQSTCMIRVLKDLKKIVDEKAIQLIHQTNP NFYVGS
SS NAAAPATVSGGNISTNQGFDDQNGMTTNPQTLLFGFDIMNRTPGV*

>G1652 (77..1078)
AGCAAGTCCAAATCTCCCTCTCTCTCTCTATCTCTCTATAGAAGATTTTTTAAAC
TAAGAAGCTAGCGATATGGCCACAGCGATGAACGTTTTCTCTACCAAATGGTCTCCGA
ATTGGATATAGAAGATATAGTATCATCCCAATTCACATGAACTCACTCGTCGGAGA
TGTTCCACAGTCTCTCTCTCTCTCTGATGATACCACCACTTGTTATAACCTTGATGCTTC
TTGTAATAAAAGTTTGGTAGAAGAAAGACCTTCAAAGATCCTCAAGACCACTCACATATC
ACCAAACCTTACATCCTTTTTCTTCTTAATCCTCCTCCTCAAAGCACCAGCCCTCTTC
TAGGATTTCTTTCTTTGAAAAGACAGGTTTACATGTTATGAATCACAACCTCTCCAAACTT
AATATTTAGCCCCAAGGACGAAGAAATTGGATTACCAGAGCATAAGAAAGCCGAGCTGAT
AATAAGAGGGACAAAGAGAGCTCAATCCTTGACTCGAAGCCAATCAAATGCTCAAGATCA
CATACTGGCAGAGAGAAAACGGAGAGAGAGAAGCTTACTCAAAGATTGTAGCTCTTCCGC
GCTAATTCCTGGCCTAAAGAAGATGGACAAGGCTTCTGTGTTGGGAGATGCAATAAAGCA
TATAAAGTACCTCCAAGAGAGTGTGAAAGAGTATGAGGAACAAAAGAAGGAAAAGACAAT
GGAATCAGTGGTTCTTGTAAAGAAGTCTAGTCTGTTTTAGATGAAAATCATCAACCATC
ATCATCATCTTCCTCAGATGGAAATCGCAATAGCTCGAGCTCAAATCTTCAGAAATAGA
AGTTAGGTTTCAGGAAAAGATGTTCTTATTAAGATCCTATGCGAGAAGCAAAAGGGTAA
TGTGATCAAGATTATGGGGGAGATTGAAAAGCTTGGTTTGTCTATCAACACAGCAATGT
CTTGCCCTTTGGACCACTTTTGACATCTCTATTATCGCTCAGAAGAATAACAATTTTGA
TATGAAAATCGAGGATGTTGTGAAGAACTTGAGTTTGGCTTATCAAAGCTCACTTAATT
GGTTTCACGTACATACATATACACATTCATCATCGATTTCTCCGATCGAAGAATCCAAA
ATCAGTTTTTCCATGAAAGTGGTTTTTTAGTTGTTAAGTTTGTGTATGGAGATCTTAA
GTCATTTAAAGATCCTTGTCTTGTGTTGTTAAGTGTGCTTTAAGATGCATATCATCAA
TGTTTAGTAATTTATTTCTCCAGTTTCATTTGGGACGGAATTTTTTTCGAGTTGTTGG
ATATATATTTCTCGCATGTAAAGCATTTCGTTAGTTTAAATAAACGTCGGATATGTTTCT
TTGAAAA

>G1652 Amino Acid Sequence (domain in AA coordinates: 143-215)

MATAMNVFSTKWSSELDIEEYSIIHQFHMNSLVGDVPSLSLDDTTTCYNLDASCNKS
VEERPSKILKTHISPNIHPFSSSNPPPKHQPSRILSFEXTGLHVMNHNPNLIFSPK
DEEIGLPEHKKAEIIIRGTRKRAQSLTRSQSNADHILAERKRREKLTQRFVALSALIPGL
KKMDKASVLGDAIKHIKYLQESVKEYEEQKKEKTMESVVLVKSSSLVDENHQPSSSSS
DGNRNSSSSNLPEIEVRVSGKDVLIKILCEKQKGNVIKIMGEIEKLGLSITNSNVLFPFGP
TFDISIIAQKNNNPFMDKIEDVVKNLSFGLSKLT*

>G1655 (132..755)

TTTCTAACTAGTCACATTGAGAGAGAGAGAGAGAGAGAGAAAGAGAGACTCTCAGAATCTGAAG
AAGAAGAAGAGATTGTTGTTTTGCGCTTTTATCATCGGTTTCTTTGAATCTCTGGTTTAA
AATCGGATTTAATGGTGGAGTCTCTGTTCCCGAGCATCGAAAACACAGGTGAATCGTCTC
GAAGAAAGAAGCCGAGGATATCAGAGACGGCGGAGCGGAGATAGAGGCACGACGTGTCA
ACGAAGAAAGCTTGAAGAGATGGAACGAATCGTGTGCAACAGATCTACGCTTGTAAAGC
TCGTGCAAGCTTTACGCCGAGTTTCGTGAGAGATCTCCACCACCAGCAACAACGAGACCG
ATAAATCGTCTCCGGCGCGGCGAGGAGATACGTGATACGGCGGATCGAGTTCTAGCTG
CGTCCGCTCGTGGTACGACTCGGTGGAGCAGAGCGATTTTAGCGAGTCGCGTCCGAGCGA
AGCTGAAGAAACATAGAAAGGCGAAAAAGTCAACGGGAAATTGTAAATCGAGAAAAGGTC
TCACGGAGACGAATCGGATTAAGTTACCGGCGGTTGAGAGAAAAGTGAAGATTCTTGCC
GTTTGGTTCTGGTTGCCGGAAGTCTCTGTACCGAATCTTTTAGATGAAGCGACCGATT
ACATCGCAGCGTTAGAGATGCAGGTTGAGGCCATGGAGGCTCTCGCCGAACTTTTAACCG
CAGCCGCACACGACGACGTTGACCGGAACCTAACGGCGGAGTTAGTTTGTAGTGTGT
TAATTAGCTTTTCTTTTACCTTTTACCCTTTTATTTTGGCTTCAAGTGTTTTTTTTTTC
TCGTGACGCGATTTTAATTTATTAAATTC

>G1655 Amino Acid Sequence (domain in AA coordinates: 134-192)

MVESLFPSIENTGESSRRKKPRISETAEAEIEARRVNEESLKRWKTRVQQIYACKLVEA
LRRVRQRSSTTSNNETDKLVSGAAREIRDADRVLAAARGTTRWSRAILASRVRAKLKK
HRKAKKSTGNCXSRKGLTETNRIKLPAVERKLKILGRLVPGCRKVSVPNLLDEATDYIAA
LEMQVRAMEALAEELLTAAAPRTTLTGT*

>G1671 (188..751)

TCCCACTATCCTTCGCAAGACCCTTCTCTATATAAGGAAGTTCATTTCAATTTGGAGAGG
ACACGCTGACAAGCTGACTCTAGCAGATCTGGTACCGTCGACCCTCTCTATATAATCTTC
TTCTACACACACACACACGCAACCATATACGTACATGTGAAGTAGTGAGATCAATATC
GTTAGCAATGAATCTACCAACGGGATTTAGGTTTTTTCCGACCGATGAAGAGCTCGTCGT
TCACCTTCTCCACCGGAAAGCTTCCCTCTTGCTTGTACCCCTGATGTATCCCCGACCT
TGATCTTTACCATTACGATCCTTGGGACCTTCCCGGAAAGCTTTGGGAGAAGGGAGGCA
ATGGTACTTCTATAGTAGAAAGACACAAGAGAGAGTGACAAGCAATGGGTATTGGGGATC
AATGGGAATGGACGAGCCAATCTACACAAGCTCCACACACAAGAAAGTGGGAATCAAAA
GTATCTAACTTTCTATCTCGGAGATTCTCAGACTAATTGGATCATGCAAGAAATATCCCT
CCCGGATTCTCTTCTCATCTAGTCGATCTTCTAAGAGATCAAGCCGTGCTTCTAGTTC
TAGTCACAAACCGATTATAGCAAGTGGGTGATATGCAGAGTGTATGAGCAAAATTGCAG
TGAGGAGGAAGACGATGATGGGACAGAACTCTCATGTTTGGATGAAGTGTTTTTGTCTTT
AGATGATCTTGACGAAGTAAGCTTACCGTAATAAAGACAGAAGCACCCAAGAAGAGAAAA
AAAAAAAAGGGTTTAGTGGGCAATTATTTCTAAGCGACCGCTCTAGACAGGCTAGTAC
CGGATCCTCTAGCTAGAGCTTTCGTTTCGTATCATCGGTTTTCGACAACGTTTCGTCAAGT

>G1671 Amino Acid Sequence (domain in AA coordinates: TBD)

MNLPPGFRFFPTDEELVVHFLHRKASLLPCHPDVIPDLDLHYHDPWDLPGKALGEGRQWY
FYSRKTQERVTSNGYWGSMGMDPIYTSSTHKKVGIKKYLTFLGDSQTNWIMQEYSLPD
SSSSSRSSSRSSSRSSSKPDYSKWVICRVYEQNCSEEDDDGTELSCLDEVFLSLDD
LDEVSLP*

>G1756 (71..1003)

ATATGTACTTGTACACCAACCCACCAAAAGAGATAAAAGAGGAAACAAAACCTCGAAAAG
AGAGAGATATATGGGTGAGGTGGCTTATATGGACGAAGGAGACCTAGAAGCAATAGTCAG
AGGCTACTCCGGCTCCGGAGACGCGTTTTCCGGCGAAAGTTCCGGTACGTTTTTCACCTTC
GTTTTGCCTACCGATGGAGACGTCTAGTTTCTACGAACCGGAGATGGAGACAAGTGGCTT
AGATGAGCTCGGTGAACTTTACAAACCTTTTACCCTTTCTCCACACAAACGATCCTCAC
AAGCTCGGTCTCTCTCCCTGAAGATTCAAAACCTTTCCGAGATGACAAGAAACAACGATC
ACATGGTTGTCTTTTATCCAACGGATCAAGAGCTGATCATATCCGAATTTCAGAATCCAA
ATCAAAGAAAAGCAAGAAGAATCAACAGAAGAGAGTTGTTGAGCAAGTGAAAGAAGAGAA

TCTGTTGTCGGACGCATGGGCGTGGCGTAAATACGGGCAGAAACCCATCAAAGGATCTCC
ATACCCAAGGAGTTATTACAGATGCAGTAGCTCAAAGGGTGTGTCAGCAAGAAAACAAGT
CGAAAGAAATCCTCAAACCCGGAGAAATTCACCATAACATACACTAATGAGCACAAATCA
TGAAC TACCAACCCGGAGAACTCATTAGCCGGTTCGACTCGAGCAAAAACCTCCCAACC
CAAACCAACCTTAACCAAAAAATCCGAAAAAGAGTTGTTTCTTCCCCTACAAGTAATCC
TATGATCCCATCCGCTGATGAATCTTCTGTTGCGGTTCAAGAAATGAGCGTTGCGGAAAC
GAGTACGCACCAAGCGGCTGGAGCAATCGAGGGCCGCGCTTGAGTAACGGTTTACCATC
GGATTTGATGTCCGGGAGCGGAAC TTTTCCAAGTTTACCAGGTGACTTCGATGAAC TATT
GAATAGCCAAGAGTTCTTCAGTGGGTATTTATGGAATTACTAGAGAGCATTAGGTGTATG
TATATATATAT

>G1756 Amino Acid Sequence (domain in AA coordinates: TBD)
MGEVAYMDEGDLEAIVRGYS GSGDAFSGESSGTFSPSFCLPMETSSFYEP EMETSGLDL
GELYKPFYFPSTQTILTSSVSLPEDSKPFRDDKKQRSHGCLLSNGSRADHIRISESKSKK
SKKNQQRVVEQVKEENLLSDAWARKYQKPKIKGSPYPRSYRCSKGLARKQVERN
PONPEKFTITYTNEHNHELPTRRNSLAGSTRAKTSQPKPTLTKKSEKEVVSSPTSNPMIP
SADESSVAVQEMSVAETSTHQAAGAI EGRRLSNGLPSDLMSGSGTFPSFTGDFDELLNSQ
EFFSGYLWNY*

>G1757 (250..1224)

ATCACCAATCCTATAACACTCTCATTCTCATCATATCATTCTTCAATCTATATAACCCAT
TCTTAATTATACTCAACACACATTATATTTTCTGATCATATCATTCTTTCAGTCCATCT
ATATAACCAATCTCTGATTTTACTTAAACACACATTATACATCTTCTCATCATAGTT
TGTATCAATTTCTAGAGTAACTACCTAAAGGAAAAAAATCTATTTGGGAATCAT
ATACTAAAAATGGAAGGAAGAGATATGTTAAGTTGGGAGCAAAAGACATTGCTAAGCGAG
CTTATCAATGGATTGATGCGGCCAAAAAGCTTCAGGCACGACTTAGAGAAGCTCCGTCG
CCGTCGTCATCATTTTTCATCACC GCGACGCTGTTGCTGAGACTAACGAGATTCTGGTG
AAGCAGATAGTTTCTTCTACGAGAGATCTCTTCTTCTGCTAAACTGGTCATCTCACC
AGCGTACAACCTTATCCGACGCCGTTACTGTAGTCCCGGTGGCAAATCCCGGCAGTGTT
CCAGAATCTCCGGCATCGATAAACGGAAGTCCGAGAAGTGAAGAGTTTGCCGATGGAGGA
GGTTCTAGCGAGAGTCATCATCGCCAAGATTACATTTTCAATTCAAAGAAAAGAAAGATG
TTACCAAAGTGGTCAGAAAAAGTGAGAATAAGCCAGAGAGAGGCTTAGAAGGACCTCAA
GATGATGTCTTTAGCTGGAGAAAATATGGTCAAAAAGACATTTTAGGCGCAAATTTCCA
AGGAGTTATTACAGATGCACACATCGTAGCACACAAACTGTTGGGCAACGAAACAAGTC
CAGAGATCAGACGGGGATGCTACGGTTTTCGAAGTGACGTACAGAGGAACACACACTTGT
TCGCAGGCGATCACAAGAACACACCATTAGCCTCGCCGGAGAGCGACAAGACACCAGA
GTCAAACCAGCCATTACCCAAAAGCCAAAGGATATTCTCGAGAGTCTTAAATCCAACCTTA
ACCGTTTGAACCGATGGGCTTGATGATGGTAAAGACGTTTTCTCGTTCCCTGATACGCCG
CCGTTTTACAAATTACGGAATATCAACGGCGAGTTCCGCCACGTGGAGAGTTCTCCGATC
TTTCGACGTTGTTGACTGGTTCAATCCAACGGTCGAGATTGACACAACCTTCCCCCGGTTT
TTACACGAGTCGATTTATTATTAATTAATAATTTGTAACAGAGAAATAGATAGTAAGTAGT
AAGTAATGATCAGCGAGAGTTAAACATAAAAGTACTTAGAGTAATCTAACGATGCATAA
TAAGGAATGTTCAACAGGACTTGAACATGATTTCAATACTAAGAGAGATTTATCTAGCTA
CTGGTAGTAGCCGAGACTTCTTGTGTAGCTTCACTTNCCTTTTGTGCTT

>G1757 Amino Acid Sequence (domain in AA coordinates: 158-218)
MEGRDMLSWBQKTL LSELINGFDAAKKLQARLREAPSPSSSFSPATAVAETNEILVKQI
VSSYERSLLLLNWSSSPVQLIPTVTVVPVANPGSVPE SPASINGSRSEEFADGGGSS
ESHHRQDIYFNSKKRMLPKWSEKVRISPERGLEGPQDDVFSWRKYGQK DILGAKFPRSY
YRCTHRSTQNCWATKQVQRSDG DATVFVETRYRGTHTC SQAITRTPPLASPEKRQDTRVKP
AITQPKPDILES LKSNLTVRTDGLDDGKDVFSFPDTPPFYNYGTINGEF GHVRESSPIFDV
VDWFNPTVEIDTTFPAFLHESIYY*

>G1782 (1..927)

ATGCAAGTGTTC AAAGGAAAGAAGATTCATCTGGGGAAACTCAATGCCTACAACAAAT
TCAAATATTCAAGGATCTGAATCTTTCAGCTTGACTAAGGATATGATAATGTCTACAACA
CAATTACCCGCGATGAAACATTCGGGTTTGAGCTGCAAAATCAAGATTCAACCTCATCA
CAATCTACTGAAGAAGAATCAGGCGCGGTGAAGTTGCAAGCTTTGGAGAATATAAGCGT
TATGGATGCAGCATTTGTTAATAACAATCTCTCAGGTTACATCGAAAACCTGGGAAAGCCT
ATTGAAAATTATACTAAGTCAATTACTACCTCGTCGATGGTGTCTCAAGACTCTGTGTTT
CCTGCTCCTACTTCTGGTCAAATATCTTGGTCTCTTCAATGTGCTGAAACGTCACATTTT

AATGGTTTCTTGGCTCCTGAATATGCATCAACACCAACGGCGCTGCCACATTTAGAGATG
ATGGGTTTGGTTTCTTCAAGAGTGCCATTGCCTCATCACATTCAAGAGAATGAACCAATA
TTTGTCAATGCGAAACAGTATCATGCGATTCTCCGTCGCAGGAAGCACCGTGCTAAACTC
GAAGCTCAGAACAACTCATCAAATGCCGTAAACCGTACCTTCATGAGTCTCGCCATCTT
CATGCTTTAAAGAGAGCTAGAGGCTCCGGTGGACGTTTCTCAATACAAAGAAGCTTCAA
GAATCATCAAACCTACTGTGTTCTTCTCAAATGGCAAATGGACAAAATTTCTCTATGAGC
CCTCACGGTGGTGGGAAGCGGAATCGGGTCTAGTTCGATCTCACCGAGCTCCAATTCAAAC
TGTATCAACATGTTCCAAAACCCGCAGTTCAGATTCTCAGGTTATCCGTCAACACACCAT
GCCTCAGCTCTCATGTCAGGGACTTGA

>G1782 Amino Acid Sequence (domain in AA coordinates: 166-238)
MQVFQRKEDSSWGNMPTTNSNIQGSSEFSLTKDMIMSTTQLPAMKHSLQLQNDSTSS
QSTBEESGGGEVASFGEYKRYGCSIVNNNLSGYIENLGKPIENYTKSITTSSMVSQDSVF
PAPTSGQISWSLQCAETSHFNGFLAPEYASTPTALPHLEMMGLVSSRVPLPHHIQENEPI
FVNAKQYHAILRRRKHRKRALEAQNKLKCRKPYLHESRHLHALKRARGSGGRFLNTKKLQ
ESSNSLCSSQMANGQNFMSPHGGGSGIGSSSISPSNSNCINMFQNPQFRFSGYPSTHH
ASALMSGT*

>G184 (327..1937)

TGAATTCTAGCCTTTTTTGTAGGCGAATCATCTGGACCGTAAGAGACTCTCTCATCGATA
ATAACCACATAATTTAATCAAACCTTTTCTCTCTTTCTAAGATCTTTTGCTTTGCTCT
TTTCTTTTTTGATCTTCTATATATGGAGAAGCACAAAACGGTACTTACTATACGATAC
TGTACGGATCCATCAAACCTGGATTAATTATCAAACCGTACATTTTTATCTTACCTGGCAA
GTTACATTCTAGGGTTTTTGGAGAATCCAATCAACAACAAAGAAAATAATCATCGTTACA
ATAATCAGTATCACGCACAGACTTAGATGTTCCGGTTTTCCAGTGAGTCTAGGCGGTTTAC
GTGACGAAGACCGTCACGATCAGATCACACCGTTGGATGACCATCGTGTGGTGGTTGATG
AGGTTGACTTCTTCTCAGAGAAGAGAGATAGGGTTTACGTTGAGAATCAACGACGACG
ACGACGAAGGCAATAAGGTTCTCATCAAATGGAGGGTTTACGAGTTGAAGAAAACGATC
GTTCCAGAGATGTCAATATCGGTCTGAATCTTCTGACCGCAATACGGGAAGCGATGAGT
CAACGGTGGATGATGGACTATCAATGGATATGGAAGATAACCGTGCAAAGATTGAGAACG
CACAACCTACAAGAAGAGCTCAAGAAGATGAAAATAGAGAATCAAAGGCTAAGAGATATGT
TGAGCCAAGCGACGACCAACTTCAATGCCTTACAATGCAACTTGTGTCCTCATGAGGC
AACGAAGAACAACGTAACCTTTCACAAGATCATCTCTGGAGAGCAAAGCAGAAGGAAGGA
AACGGCAGGAACCTGCAAAATCATGGTGGCCAAGGCAGTTTATGGACCTTGGGCCGTCGTCTG
GAGCAGCAGAGCATGGAGCCGAAGTGTCTATCTGAAGAGAGGACAACGGTTCGTTACAGTT
CTCTCTCTTCGCTTCTAGAAAGTTCCAATCCCCGAGAGAACGGAAGAGGTTGCTTGGAA
GAGAAGAAAGCTCAGAGGAATCAGAGTCTAACGCCTGGGGAAACCCCTAACAAAGTCCCCA
AACATAATCCATCCTCTAGCAATAGCAATGGAACAGAAACGGAATGTTATTGATCAGT
CGGCCGAGAAAGCCACCATGCGGAAAGCCCGTGTCTCAGTTCTGTCCTGATCTGAAGCTG
CCATGATAAGCGATGGATGTCAATGGAGAAAGTACGGACAAAAAATGGCTAAAGGAAACC
CGTGTCCCGGGCTTATTATCGTTGCACAAATGGCCGGTGGATGTCAGTTTCGCAAGCAAG
TGCAGCGTTGCGCAGAAAGACAGATCTATTCTCATAACCACCTACGAAGGAAACCACAACC
ATCCACTCCACACGCGCTACGGCCATGGCCTCAACAACCACCGCAGCTGCAAGCATGC
TCCTCTCGGGCTCAATGTCGAGTCAAGACGGTTTAAATGAACCCAAACAACTCTCTAGCTC
GAGCTATCTTGCCTTGCTCTCAAGCATGGCTACAATCTCAGCTCCGCACCATTTCCAA
CCATCACATTGGACCTCACCAATTCACCCAACGGTAACAACCTAATATGACCACTAATA
ACCGGTTGATGCAGTTTCGCTCAACGGCCCGGTTTCAACCCGGCAGTTTTCCTCAAGTGG
TTGGTCAAGCTATGTACAAATAACCAACAACAGTCCAAGTTTTCTGGTTTACAGTTACCGG
CTCAGCCACTGCAGATCGCGGCCACTTCTCGGTGGCCGAGAGCGTTAGTGCTGCCAGTG
CAGCAATTGCGTCCSATCCAACTTTGCGGCGGCTCTAGCGGCAGCGATCACGTCCATTA
TGAACGGTTCCAGTCATCAAATAATAACACCAATAATAATGTGGCTACGAGCAACA
ATGACAGTAGGCAATAAGAGTTTTTCAATTTGATGGTTCGATTTTTTTTTTTTGGGG

>G184 Amino Acid Sequence (domain in AA coordinates: 295-352)
MFRFPVSLGGSRDEDHRDQITPLDDHRVVDEVDFSEKRDVRSRENINDDDDEGNKVL
KMEGSRVEENDRSRDVNI GLNLLTANTGSDESTVDDGLSMDMEDKRAKIENALQLEELKK
MKIENQRLRLDMLSQATTNFNALQMLVAVMRQEQNRNSSQDHLLESKAEGRKQELQIMV
PRQFMDLGPSSGAAEHGAEVSSERTTVRSGPSPLLESSNPRENKRLLGREESSESE
SNAWGNPNKVPKHNPSNSSNNGNRNGNVIDQSAABATMRKARVSVRARSEAAMISDGCQW
RKYGQKMAKGNPCPRAYYRCTMAGGCPVRKQVQRCAEDRSILITTYEGNHNHPLPPAATA

MASTTTAAASMLLSGSMSSQDGLMNPTNLLARAILPCSSSMATISASAPFPPTITLDTNS
PNGNNPNMTTNNPLMQFAQRPGFNPAVLQVVGQAMYNQKSKFSGQLQPAQPLQIAAT
SSVAESVSAASAAIASDPNFAAALAAITSIMNGSSHQNNNTNNNNVATSNNDSTRQ*
>G1845 (111..989)
AAGACATAATTTCTCTGTTTTCTCTAGCTCTCTCTCTCAAAATCTTCCATTGCTCTCTG
TTTTGGCAAAATCGTGAAGTCCACGCTTTAAGGCATCAGTGAAGCAAAGATGGACTTTG
ACGAGGAGCTAAATCTTTGTATTACGAAAGGTAAAAATGTTGATCATTCTTTTGGAGGAG
AAGCTTCTTCCACGTCCCAAGATCTATGAAGAAAATGAAGAGTCTAGTCGTCCTAAAC
CCTATTTCCAATCCTCTCTCTCTCTCTTATTCTGTTAGAGGCTTTCCCTTTTTCTCTCGATC
CAACACTTCAGAATCAGCAACAACAACCTCGGATCATACGTTCCGGTACTTGAGCAACGAC
AAGACCCGACAATGCAGGCCAGAAGCAAATGATCTCTTTAGTCCTCAACAACAACAAC
AGCAGCAGCAGTATATGGCCAGTACTGGAGTGACACATTGAATCTGAGTCCAAGAGGAA
GAATGATGATGATGATGAGCCAAGAAGCTGTTCAACCTTACATCGCAACGAAGCTGTACA
GAGGAGTGAGACAACGCTTAAATGGGTGCGAGATCCGTAAGCCACGAAGCA
GGGCACGCTTTTGGCTTGGTACCTTTGATACAGCTGAAGAAGCTGCCATGGCCTACGACC
GCCAAGCCTTCAAATTACGAGGCCACAGCGCAACACTGAATTTCCCGAGCATTTTGTGA
ATAAGGAAAGCGAGCTGCATGATTCAAACCTCGTCGGATCAGAAAGAACTGAAACGCCAC
AGCCAAGCGAGGTTAACTTGGAGAGCAAGGAACCTACCGGTGATTGATGTTGGGAGAGAGG
AAGGTATGGCTGAGGCATGGTACAATGCCATTACATCGGGATGGGGTCTGAAAGTCCCTC
TTTGGGATGATTGGATAGTTCTCATCAGTTTTTCATCAGAAAGCTCATCTTCTCTCTCTC
TCTCTTGTCTATGAGGCCTTTCTTTTGAAAAAGTTTATAAACCCACATTGTGTTGTAGG
TTATAGTTTAGGGTTATGCTCATTGGCATTGGGATGGAGGCAATTTTGTGATCTCCCAT
TCCACCACATATCAGTCATTATATGTGTCTACCTTTTCTCTGTATTTCTATCATTATCAT
TGTTTTTATTATGTCTGTATGTGTTTCCCTATTGCTACATACATAGATGTCCTCTTTG
TTCAAAAA
>G1845 Amino Acid Sequence (domain in AA coordinates: 140-207)
MDFDEELNLCITKGNVDHSGGEASSTSPRSMKMKSPSRPKPYFQSSSSPYSLEAFPF
SLDPTLQNOQQQLGSYVPLEQRQDPTMQGQKQMSFSPPQQQQQQQYMAQYWSDTLNL
PRGRMMMSQEAQVQPIATKLYRGVRQRQWGWVABIRKPRSRARLWLTFFDTAEAAAM
AYDRQAFKLGRHSATLNFPEHFVNKESELHDSNSSDQKEPETPQPSEVNLESKELPVIDV
GREGMAEAWYNATTSWGPESPLWDDLDSSHQFSSSSSSPLSCPMRPF*
>G1879 (3..917)
AAATGCCCTTAGAGGCTGTCTATACCCGCAAGATCCATTCCGATATCTCTCCAATTGCA
AAGATTTTATGTTCCACGACTTATACTCTCAAGAAGAGTTCGTAGCTCAAGATACGAAGA
ACAACATTGATAAGTTAGGCATGAACAGAGCTTTGTGGAACAAGGTAAGGAGGACGATC
ATCAATGGCGAGACTATCATCAGTATCCTTTGTTGATCCCTTCGTTGGGAGAAGAGCTTG
GTCTTACCGCCATTGATGTGAGAGTCACTCTCTCCACAGCACCGGAGGAAGAGGAGGA
GAACGAGAACTGCAAGAACAAAGGAAGAGATCGAGAACCAGAGAATGACTCACATCGCCG
TCGAGAGAAATCGCCGAAACAGATGAACGAGTATCTGGCTGTGCTCCGTTCTCTAATGC
CGTCGTCTGATGCTCAAAGAGGAGATCAAGCGTCGATAGTAGGAGGAGCTATAAACTACG
TGAAGGAGTTAGAGCATATTTTACAATCTATGGAGCGGAAGAGAACTAGGACTCATGATC
CCAAAGGAGACAAGACTAGCACTAGCTCGTTAGTGGGTCCATTACAGATTTTTTTCAGCT
TCCCACAATATTCTACAAAGTCAATCATCAGATGTACCGGAAAGCTCATCTTACCGGCGG
AGATAGAGGTTACGGTGGCAGAAAGCCATGCGAACATCAAGATAATGACGAAGAAGAAAC
CGAGGCAGCTTCTTAAGCTCATACTTCTTTACAAAGCCTAAGGCTCACTCTTCTTCATC
TCAATGTCACTCTCCCAACTCCATTCTCTACTCCATCAGCGTCAGGGTTGAAGAAG
GAAGCCAACCTGAATACCGTGGACGACATTGCAACAGCTTTGAATCAAACATAAGGAGGA
TTCAAGAAGAGACATAATTGAGCAAAATAGATTATAATTAACCTGTTTATTTTATTTTA
TTTTGAAATAACTGAAATCAGTTTTCTAATTTTTTTTTTTTTTCACTATTCTCTAATCC
TCCCTATGTAGTTGATTTTTGTCTCTTGTAAATGAATCAATGGTCATAAAGATCTGAAC
AAAAAATTGAATAAAGAAAATGGTT
>G1879 Amino Acid Sequence (domain in AA coordinates: 107-176)
MPLEAVVYPQDPFGYLSNCKDFMFHDLYSQEEFVAQDTKNNIDKLGHEQSFEQGKEDDH
QWRDYHQYPLLIPSLGELGLTAIDVESHPPQHRKRRTNRNCKNKEIENQRMTHIAV
ERNRRKQMNELAVLRSLMPSSYAQRGDQASIVGGAINVVKLEHILQSMPEKRTTRTHDP
KGDKTSTSLVGPPTDFSFPPQYSTKSSSDVPSSSSPAEIEVTVAESHANIKIMTKKPK
RQLLKLITSLQSLRLTLHLNVTTLHNSILYSISVRVEEGSQLNTVDDIATALNQTIRRI

QEET*

>G1888 (1..729)

ATGAAGATTGGTGTGCTGTTTGTGATAAAGAAGAAGCTTCGGTGTGTTTGTGTGCGGAT
GAAGCAGCTCTTTGTAATGGTTGCGATCGCCATGTTCAATTCGCCAATAAACTAGCCGGG
AAACATCTCCGGTTCTCTCTCACTTCTCCTACTTTCAAAGATGCTCCTCTTTGTGATATT
TGCGGGGAGAGGCGTGCATTATTATTTTCCAAGAAGACAGAGCAATACTATGCAGAGAA
TGTGACATTCCAATACATCAAGCTAATGAGCACACTAAGAAACACAATAGATTCTCCTCT
ACCGGCGTTAAGATCTCTGCTCCCGTCAGCCTACCCAAGAGCCTCCAATTCCTCACTCT
GCTGCTGCATTTGGTCGAGCCAAAACCCGACCAAAATCAGTATCGAGCGAGGTCCCAGC
TCGGCCTCCAATGAGGTATTTACGAGCTCTTCTTCGACGACCACGAGCAATTGCTATTAT
GGGATAGAAGAAAACCTACCATCACGTGAGCGATTTCGGGGTCGGGATCGGGTTGTACAGGT
AGTATATCCGAGTATTTGATGGAGACATTACCGGGTTGGAGAGTGGAGGATTGTCTTGAA
CACCCTTCTTGTGTCTCCTATGAGGATAACATTATTACTAATAACAATAACAGTGAGTCT
TATAGGGTTTATGATGGTTCTTCACAATTCATCATCAAGGGTTTGGGATCACAACCC
TTCTCTTGA

>G1888 Amino Acid Sequence (domain in aa coordinates: 5-50)

MKIWCAVCDKEEASVFCCADEAALCNGCDRHVHFANKLAGKHLRFSLSPTFKDAPLCDI
CGERRALLFCQEDRAILCRECDIPIHQANEHTKKHNRFLTGVKISASPSAYPRASNSNS
AAAFGRAKTRPKSVSSEVPSSASNEVFTSSSSTTTSNICYGIEENYHHVSDSGSGSGCTG
SISEYLMETLPGWRVEDLLEHPSVSYEDNIITNNNNSES YRVYDGSSQFHHQGFWDHKP
FS*

>G189 (34..987)

CCACAACCTCTCTCCTTGTAGAGAGAGAGATTTTATGGCGGTGGAGCTCATGACTCGGAAT
TACATCTCCGGCGTCGGAGCTGATAGCTTCGCGGTTCAAGAAGCAGCTGCTTCAGGACTC
AAAAGTATCGAAAATTTATCGGTTTAATGTCTCGTGATAGCTTAACTCTGATCAGCCA
TCTTCTTCTTCCGCCCTCCGCCCTCCGCCCGCCGAGATCTGAATCAGCTCGTAAC
ACAACGGCGGACGCGGCTGTTTCAAAGTTTAAAAGAGTCATATCTCTCTTAGATCGAACT
CGAACCGGACACGCCCCGTTTAGACGTGCTCCGGTTCATGTTATTTCTCCGGTTCTTTTA
CAAGAAGAACCAAAAACGACGCCGTTTCAGTCTCCTCTCTCCTCCGCCGCAATGATC
CGAAAAGGTTTCGTTTTCTTCATCGATGAAAACGATTGATTTCTCATCTCTCTCTCTGTA
ACAACCGAATCAGACAACCAGAAGAAGATTATCATCATCAACGTCCTCTGAAACGGCG
CCGTTTTCGCTCTCAAACTCAAAGCCTCTCCACGACGGTCTCGTCTTTCTCAAAATCAACA
AAGAGAAAATGTAACTCTGAGAATCTTCTCACCGGAAAATGCGCTTCCGCTTCTCTCTCC
GGTCGTTGTCTATTGCTCGAAGAAAAGAAAGATAAAACAGAGGAGAATAATTAGGGTTCCG
GCGATAAGTGCAAAAATGTCCGATGTACCACCGACGATTATTATGGAGGAAATACGGA
CAAAAACCAATTAAGGATCTCCACATCCAAGAGGATATTATAAGTGTAGTAGCGTAAGA
GGTTGTCCAGCACGTAACATGTTGAGAGAGCAGCTGATGATTCTGTCATGTTGATTGTT
ACTTATGAAGGAGATCATAATCATTCTCTCTCCGCCGCTGATCTCGCCGAGCCGCCGTT
GCTGATCTTATTTTGAATCGTCTTGAAAAGAACAATCTTTATTTAAGGCTTTTATAAT
ATAAATTTAGATCCTTACTTAGTGAAGTACTCAAATATGAATGAAATCAATGTAATCAA
AATCAAAAAGCTTTTGTCTAAAAA

>G189 Amino Acid Sequence (domain in AA coordinates: 240-297)

MAVELMTRNYISGVGADSFVQEAASGLKSIENFIGLMSRDSFNSDQPSSSASASASA
AADLESARNTTADA AVSKFRVISLLDRTRTG HARFRRAPVHVISPVLQEEPKTTPFQS
PLPPPPQMIRKGSFSSSMKTIDFSSLSSVTTESDNQKKIHHHQRPSETAPFASQTQSLST
TVSSFSKSTRKCNSENLLTGKCAASSSGRCHCSKKRKIKQRRRIIRVPAISAKMSDVPP
DDYSWRKYGQKPIKGSPPHGRGYKCSSVRGCPARKHVERAADSSMLIVTYEGDHNHSL
AADLAGAAVADLILESS*

>G1939 (92..844)

AATCATTAGCTTCTTCTCTTCTCTCTCACAGAGAGAGTAATCACAAGCCAAGTGAGA
AAAAGAAAACACTAAACCCAGATCGAAAACCATGTCTATTAACAACAACAACAACAA
CAACAATAACAACGATGGTCTTATGATCTCATCAAACGGAGCTTTAATCGAACAACAAC
ATCAGTCGTTGTGAAGAAACACCGGCGAAAGATCGACATAGCAAAGTCGATGGAAGAGG
GAGAAGAATCCGTATGCCGATTATATGTGCTGCTCGTGTGTTTTTTCAGCTAACGAGAGCT
TGGTCATAAGTCAGATGGCCAAACAATTGAATGGTTACTTTCGTCAGCAGAGCCTTCTAT
TATAGCTGCAACAGGAACGGTACAACCTCCAGCGAGTTTCTCAACTGCTTCTGTCTCTAT
CCGTGGAGCCACCAATCTACTTCTTTAGATCATAAACCCACTTCTTTACTTGGTGGTAC

GTCACCGTTTATACTTGGGAAACGTGT'TAGAGCTGATGAGGATAGTAATAATAGTCATAA
TCATAGTTCTGTTGGTAAAGATGAGACCTTTACGACAACACCAGCTGGGTTTTGGGCTGT
TCCGGCGAGGCCGGGATTTTGGACAAGTTTGGAGTTTGTCTGGAGCTCCAÇAAGAGATGTT
TTTACAACAACAACATCATCATCAGCAACCATTGTTTGTTCATCAGCAACAGCAACAACA
AGCTGCAATGGGTGAAGCTTCTGCTGCTAGAGTTGGGAATTATCTTCCGGGTCATCTTAA
TTTGTCTTGTCTTCTTTATCCGGTGGATCTCCCGGGTCGGATCGAAGAGAGGAAGATCCACG
TTAATGGTTTAAAGCCCTTTTAGGTTTGAAGGCAAAATTTGGTATATATATTTATTATCTT
CTCTTCTCTATTGTTGTCTATTGTTTCTCTATGTGTGTGTTTTAGTGTGTTAGAGATTGA
TTTGGTTTTCAGAATCTCTGCAAGTGATTGAGAGTTTTCGTTAGCTTTAAGTAAGTTAAA
GACGGTTGTTTTGATTAGGGTTAAATTAGGGTTTAAAGAATCTGTTGTTTTTTTGGAGGG
AGATCGATTCTTATCGGATCCAAGATTACTTTTAGGAAAAAAGGAAAAATTCAGAAAC
CACGGTGGTTTCTTTTCTCTTTTTTTTTTTT

>G1939 Amino Acid Sequence (domain in AA coordinates: 40-102)
MSINNNNNNNNNNDGLMISSNGALIEQQPSVVVKPPAKDRHSKVDGRGRRIRMPIICA
ARVFQLTRELGHSKSDGQTIEWLLRQAEPSIIAATGTGTPASFSTASVSIRGATNSTSLD
HKPTSLGGTSPFILKRVRAEDSNNSHNHSSVGKDETF'TTPAGFWAVPARPDFGQVW
SFAGAPQEMFLQQQHHHQQPLFVHQQQQQAAMGEASAARVGNYPGHLNLLASLSGGSP
GSDRREEDPR*

>G194 (192..1205)
TCTTCTCTCTCTCTCTCTCTCTCTTGAACCCCTAAAACTCTTTCTTTTACAAGGATT
GATCTTTTTGTATT'TTTGATTTTGTGACATTTGCTTTGTGTTTCGATCTCTGTTTTGATGCGA
TTTCTCTGTTTTTAAAGCCATTGATAGATTGTTTCCGGTAAAGCTCAGCGAGAGAAGAA
GAAGAACAACATGGAGTTTACAGATTCTCAAAGACGAGTTTTTACTACCCGTCGTCAC
AAAGCGTTTGGGATTTTCGGAGATTAGCGGCGGCGGAGAGGCATTCTTTAGGGTTCATGG
AGTTATTAAGTTCTCAGCAGCATCAAGACTTTGCTACTGTTTCTCCTCATTCCTTCCTTC
TCCAAACGTCTCAACCGCAAACGCAACCAACCATCGGCGAAGCTGTCTTCAAGTATCA
TTCAAGCTCCACCGTCAGAGCAATTAGTGACGTCAAAGGTGGAGTCTTTGTGTTCCGATC
ATTTGTGTGATAAACCACCGCGACTCCTAACTCGTCATCGATTTCGTCTGCTTCAAGCG
AGGCTCTAAATGAAGAGAAACCGAAAAAGAGACAATGAAGAAGAAGGAGGTGAAGATC
AACAAGAGAAGAGTCATACTAAGAAACAGTTGAAAGCAAAGAAGAATAATCAGAAGAGAC
AGAGAGAGGCAAGAGTCGCATTATGACAAAGAGTGAAGTTGATCATCTCGAAGATGGTT
ATCGCTGGCGAAAAATATGGTCAAAAAGCTGTCAAAAACAGTCTTTTCCAGGAGTTACT
ACCGTTGCACAACGGCTTCATGTAACGTGAAGAAGAGAGTGGAGAGATCATTCAGAGATC
CAAGCACTGTGGTTACAACCTACGAAGGTCAACACACTCACATTAGTCCACTCACGTCTC
GTCCTATTTCCACTGGAGGTTTCTTCGGATCGTCAGGAGCTGCTTCGAGTCTCGGTAATG
GTTGCTTTGGGTTTCTATTGATGGCTCCACGTTAATCTCTCCTCAGTTCCAACAGCTTG
TCCAATACCATCACCACAGCAGCAACAAGAACTCATGTCTTGT'TTGGAGGAGTCAACG
AGTACCTTAATAGCCACGCTAATGAGTATGGTGATGATAATCGTGTGAAGAAGAGTCGAG
TTTTGGTTAAAGATAATGGACTTCTGCAAGATGTTGTTCCGTCTCATATGTTGAAGGAAG
AGTAGTAGTATATATATAGTCTTATAGTTTAAATCTAGTTTTTTTTTGTATAATTGTCTA
AAAGAAACGGATCTTTGTTCTGATGAAGAAGATGTTTCTTATGGTTCTGAAATCGTAA
GGTAATGATGATTGTACCAAGCCGAGAAAGTACTTGTGATTTTCACCATGAATCACTAT
AAATGTAATTTTATTTACTGTGAAAAAAAAAAAAAAAAA

>G194 Amino Acid Sequence (domain in AA coordinates: 174-230)
MEFTDFSKTSFYYPSSQSVWDFGDLAAERHSLGFMELLSSQQHQDFATVSPHSFLLQTS
PQTQTQPSAKLSSSIIQAPPSEQLVTSKVESLCSHDLLINPPATPNSSSISSASSEALN
EEKPKTEDNEEEGGEDQKEKSHTKQLKAKKNNQKRQREARVAFMTKSEVDHLEDGYRWR
KYGQKAVKNSPFPRSYYRCTTASCNVKKRVERFRDPSTVVTTYEGQHTHISPLTSRPI
TGGFFGSSGAASSLNGCGFGFPIDGSTLISPQFQQLVQYHHQQQQQELMSCFGGVNEYLN
SHANEYGDDNRVKSRVLVDNGLLQDVVPSHMLKEE*

>G1943 (137..1858)
ACATTTGTTTCTAATCTCAGACATAAATAATTTTTGTTCCCGACTTCAAAACCAACGATG
ATTATATCATTCACATTCATTTCTTCTACTTCTTCTTCTCTTCTGATCTCATTTCCCT
AGAAAATCCATCTATCATGGGTGAAGATGATATAGTGGAGCTCTTATGGAAGAGTGGCCA
AGTCGTTAGAACCAGTCAAACACAGAGACCTCCTCCAATACACCACCATCTCTTCTCC
ACCACCCATTCTTCGTGGTAGCGGAAGCGGCAACGGAGAAGAAAATGCCCCGCTTCCACT
TCCACAGCCTTCACCTCCCTCCATCATCAGAATCTTTTCATTCTGGAAGACGAAATGTC

TTCTTGGCTTCACCATCTCACCCCGGCGTTACGTCCACCCCGGCTTCTTCTGTCTCCCT
GCCACCACCACCAATGCTCCGCGTGAAGATGATATAGTGGAGCTTTTATGGCAAAGCGG
CCAAGTAGTTGGAACCAACCAACACATAGACAATCCTACGATCCTCCTCCATTCTCCG
CGGCAGCGGAAGTGGCAGAGGAGAAGAAATGCTCCCTTTTACAACTCCGCTCCGCTCACCT
GCATCAGCAAAATCTCTTCAATCAAGAAGGCGAAATGTATTCTGGCTACACCATCTTA
CCGCCAAAATATTTCTGCTCAGAACTTCTCAACTCCACTCCGGCTACTCACCCGCAAAG
TTCCATCTCTCTGGCACCACGTCAGACTATCGCCACGAGAAGGGCGGAAAACCTTTATGAA
CTTCTCGTGGCTAAGAGGGAACATATTTACCGCGGCTAGAGTTGATGAAGCTGGACCGTC
GTTTTCTGGTGGTAAGAGAATCGATGCAGGTAGGCTCGAACACGACCCCCCTTCTTCTC
TGCCACTGAATCATGTGTAATACCAGCTACAGAGGGCACCGGAGTCGAGTGTGCGGAAC
TTTGGCAGCTCATGATCTTGGTCGGAAGGGAAAGGCGGTGGCGGTTGAGGCGCGCGGAAC
ACCATCTTCAGGAGTGTGCAAGGCCGAAAACAGAGCCGGTTCAGATACAACAGCAACGGA
GTGCAAGCTAAAAGCGAGAGAAGAAACCATGGAAGTGAAGAAGCTCGTGGTTCAACGTC
TAGAAAGAGATCACGAATGCAGAAATGCATAACCTCGCCGAAAGGAGAAGGAGAGAAAA
GATCAACGAGAAGATGAAGACTCTGCAACAACCTATTCTCGCTGCAACAAGGTTGAATC
TGATTCTGTTTCTACTCTGATCAGTCTACTAAAGTTTCAACCGTGGATGATGCTATCGAG
TACGTCAAATCGTTACAGAGCCAAATACAAGTATGCTCTTCAAAACAGAATGTGTTTTAA
ACCAATGGTTCAACATGGAAGAGTTTCAATGTATCTAGTTTGTGTTGAGATGATGTCGAC
GGGACAGGGTATGATGTGCGCAATGATGAATGCGGGAATACGCAACAGTTTATGCCCA
TATGGCCATGGATATGAACCGACCTCCTCCATTATACCTTTCCCGGCACATCTTTTCC
TATGCCGGCTCAAATGGCAGGTGTAGGTCCATCATATCCAGCACCGCGCTACCCTTTTCC
CAACATTAGACCTTTGACCCATCCAGAGTCCGTTTACCAAGCCCGCAGCTAACCCGGT
GTCGAACCAGCTCAGTTTCCGGCTTACATGAATCCCTATAGCCAGTTTGTGTTCCCA
CCAGTTGCAACAACCTCCTCCTCCTCATTTCAGGGTCAAACAACATCACAACCTGAGTTC
CGGGCAGGCAAGTAGTAGCAAGGAACCTGAGGATCAGGAGAACCAACCAACAGCTTAGTT
AAAGTGTGGAGCTGAAACCGATCAGTTCTTCAAGCAAATTACAACCTTTGAAGATAAACCA
GAGTTGTAACATGTAGATTTTGTCTGTTAAGTTTAAATGTAAGTACTTTTGTAGTTAATGGG
AAAGATACTGACAGGTGCAAGGTGGTCAATTTGTGCATCACGCTTAAGATTCCTCGA
TGTGGCCAGTATCTCCCTTTCTAGCATGTGAGGTCCCTACTCTCTGTTTCTACGGAGAC
CAAATGTTCCAGCTGATTAAACACACAATGACTTACCAAAAGTACACGCGGCCCATCCTCG
TCTTTATGTTCCAAGTGCAGCTGTTTGTATTGTAAGCATTTTTCTTATAATAATAAA
ACAGCTCTATCTTCGTTAAAAAAA

>G1943 Amino Acid Sequence (domain in AA coordinates: 335-406)

MGEDDIVEILLWKSQVVRSTQTRPSSNTPPSLPPPPILRGSGSGNGEENAPLPLPQSP
PLHHQNLFILEDEMSSWLHSHPGVSTSTPASSVSLPPPPNAPREDDIVELLWQSGQVVG
NQTHRQSYDPPPIILRGSGSGRGEENAPLSQPPPHLHQNLFIQEGEMYSWLHHSYRQNYF
CSELLNSTPATHPQSSISLAPRQTIATRRARENFMNFSWLRGNIFTGGRVDEAGPSFSVVR
ESMQVGSNTTPSSSATESCVIPATEGTASRVSGTLAHLGRKGKAVAVEAAGTPSSGV
CKAETEPVQIQPATESKLKAREETHGTEEARGSTSRKRSRTAEMHNLAERRRRREKINEKM
KTLQQLIPRCNKVESDSVSTLISLLKFQRMMLSSSTSNRYRAKYKALQNRMCFKPMVQH
GKSSYVSSFVEMMSTGQGMSPMMNAGNTQQFMPHMMAMDMNRPPPIPFPGTSFMPAQM
AGVGPSYPAPRYPFPIQTFDPSRVRLPSPQPNVSNQPFQPAYMNPYSQFAGPHQLQQP
PPPPFQQTTSQLSSGQASSSKEPEDQENQPTA*

>G21 (79..966)

TGTGGAGGAATATTAATACAGCCCACTTCACATCTATTTTGTGCAACCATCTCTCTAAA
GCTTCTTCTCTCATAACAATGGCAAGACAAATCAACATAGAGAGTAGTGTCTCAAGTT
ACCTTTATCTCCTCCGCGATCCCCGCGTATCTTCTCCTCCTCCATCACCGCTTCCGCC
TCATTGTCTCTTGACCTACTACATCTTCTCTTCTTCGTCATCAACAAATTCTAACTTC
ATTGAGGAAGACAACCTCTAAAAGAAAAGCATCTCGAAGATCATTGTATCGTTAGTCTCC
GTTGAAGACGATGATGATCAAAACGGTGGAGGTGGGAAACGGCGAAAGACCAACGGTGG
GATAAACATCCGACGTATAGAGGAGTGAGGATGAGGAGTTGGGGAATGGGTGTGCGAG
ATTAGAGAGCCGAGAAAGAAATCAAGAATCTGGCTCGGGAATTATCCAACGGCTGAGATG
GCAGCTCGAGCTCATGACGTAGCGGCTTTAGCCATTAAAGGTACAACGGCTTACCTCAAT
TTTCCCAAGTTAGCCGCGGAGCTTCTCGTCCAGTCACAAATTCTCTAAAGACATTCAA
GCCGCCCGCTTTTAGCGGCCGTTAACTGGCAAGATTCGGTCAACGATGTGAGTAATTCT
GAAGTGGCTGAAATAGTTGAAGCCGAGCCGAGTCCAGCCGTGGTGGCTCAGTTGTTTCT
TCGGACACAAGCACGACGACGAGCTCAGAGTCAAGAGTATTCCGAAGCTTCGTGTGCT

TCGACTTCGGCGTGACGGACAAAGACAGTGAGGAAGAGAAGCTGTTTGATTGCGCGGAT
TTGTTTACCGATGAGAATGAGATGATGATACGAAACGATGCGTTTTGCTACTACTCGTCC
ACGTGGCAGCTTTGTGGAGCCGATGCTGGGTTTCGGCTTGAAGAGCCGTTTTTCTATCT
GAATGACTAAAGTACCCCTCTCGAGAGAGCTCTCACTAACACT

>G21 Amino Acid Sequence (domain in AA coordinates: 97-164)

MARQINIESSVSQVTFISSAIPAVSSSSSITASASLSSSPTSSSSSSSTNSNFIEEDNS
KRKASRRSLSSLVSVEDDDDDQNGGGGKRRKTNGGDKHPTYRGVVRMSWGKVVSEIREPRK
KSRIWLGTPTAEMAARAHDAALAIKGTAYLNFPLKLAGELPRPVTNSPKDIQAAASLA
AVNWQDSVNDVSNSEVAEIVEAEPSRAVVAQLFSSDTSTTTTTQSQEYSEASCSTACT
DKDSEEEKLFDLPDLFTDENEMIRNDAPCYSSSTWQLCGADAGFRLEEPFFLSE*

>G2132 (42..1031)

ATTCTGTACTTAGTACCGGAGTTTAGTCCGAGAGAGACAATGATCAGTTTCAGAGAAG
AGAATCATCGATCTCACTTGATTAACAATAGTGTAATCTGTAATGATCCAGACGCCA
CCGATTCCTCTAGCGACGATGAATCTATCTCCGGCAATAATCTCGCCGTCAGATCAAAC
CAAACACCACCGAAACGTTACGTCTCAAAGATCTGTGTCCGACGCTGATCAAAGGTATG
AGAACGTTTCGAATTCTACAGGGAATAAAGCAGCCGGAACCGGAAACGTCGTCGGGTT
TCAAAGGCGTACGACGAGGCGGTGGGGAAATTTGCGGCGGAGATAAGAAATCCGTTTG
AGAAGAAGAGAAAGTGGCTTGAACGTTTCTACTGAAGAAGAAGCAGCAGAAGCTTACC
AAAAGAGTAAAGAGAGTTTGATGAACGATTGGGTTTAGTTAAACAGGAAAAGACCTAG
TAGATTTGACCAAGCCGTGCGGTGTACGTAAACAGAGAAGGAAGTTACTGAGAAGT
CGAATTGCAAAAAGGTAAATAAGAGAATTGTTACTGATCAGAAGCCATTGGTTGTGGTT
ATAACGCTGATCATGAAGAAGAGGGAGTGATTAGTAAATGTTGGAAGATCCGTTGATGA
CATCGTCAATTGCTGATATTTTGGTGATTCGGCTGTGAAGCAAATGATATTTGGTG
ATTACAATTAGTGAATTTATTTCCATTGTAGATGATTTCAAGTTTGATTTTGTGGAGA
ATGATAGAGTAGGAAAGGAGAAAACATTGGATTTAAGATTGGGGATCACACTAAAGTTA
ATCAACATGCCAAAATCGTATCGACCAATGGGGACTTATTCGTCGATGATTTACTTGATT
TTGATCCGTTGATAGATGATTTAAGTTAGAAGATTTTCTATGGATGATCTTGGATTAT
TAGGAGATCCAGAGGATGATTTTAGTTGGTTTAAATGGTACTACTGATTGGATCGATA
AGTTTTTATGAATACTTTCTTGACACGGCCAACGGTATTAGTAC

>G2132 Amino Acid Sequence (domain in AA coordinates: TBD)

MISFREENIDLNLIKTISVICNDPDATSSSDDESISGNPNRRQIKPKPKRYVSKICVP
TLIKRYENVSNSTGNKAAGNRKTSSGFKGVRRRPWGKFAAEIRNPFKKRWLGFPTTEE
EAAEAYQKSKREFDERLGLVKQEKDLVDLTKPCGVKRPKEEKEVTEKSNCKKVNKRIVTDQ
KPFCCGYNADHEEEGVISKLEDPMTSSIADIFGDSAVEANDIWDVNSVEFISIVDDF
KFDFVENDRVGKEKTFGFKIGDHTKVNQHAIVSTNGDLFVDDLDFDPLIDDFKLEDFP
MDDLGLLGDPEDDDDFSWFNGTTDWDKFL*

>G2145 (1..777)

ATGGACGTTTTTGTGTGATGGTGAATTGGAGTCTCTCTTGGGGATGTTCAACTTTGATCAA
TGTTTCATCATCTAAAGAGGAGAGACCGCGAGACGAGTTGCTTGGCCTCTCTAGCCTTTAC
AATGGTCATCTTCATCAACATCAACACCATAACAATGTCTTATCTTCTGATCATCATGCT
TTCTTGCTCCCTGATATGTTCCCATTTGGTGCAATGCCGGGAGGAAATCTCCGGCCATG
CTTGATTCTTGGGATCAAAGTCATCACCTCCAAGAAACGCTTCTCTTAAGAGGAAACTA
CTTGACGTGGAGAATCTATGCAAACTAACTCTAACTGTGACGTCACAAGACAAGAGCTT
GCGAAATCCAAGAAAAACAGAGGGTAAGCTCGGAAAGCAATACAGTTGACGAGAGCAAC
ACTAATTGGGTAGATGGTCAGAGTTTAAGCAACAGTTCAGATGATGAGAAAGCTTCGGTC
ACAAGTGTTAAAGGCAAACTAGAGCCACCAAGGGACAGCCACTGATCCTCAAAGCCTT
TATGCTCGGAAACGAAGAGAGAAGATTAACGAAAGGCTCAAGACACTACAAAACCTTGTG
CCAAACGGGACAAAAGTCGATATAAGCACGATGCTTGAAGAAGCGGTCCATTACGTGAAG
TTCTTGACGCTTCAGATTAAGTTGTTGAGCTCGGATGATCTATGGATGTACGCACCATG
GCTTACAACGGCCTGGACATGGGGTTCATCACAACCTTTTGTCTCGGCTTATGTGA

>G2145 Amino Acid Sequence (domain in AA coordinates: 166-243)

MDVFDGELLESLLGMFNFDQSSSKEERPRDELLGLSSLYNHLHQHQNHNVLSSDHHA
FLLPDMFFPGAMPGGNLPAMLDSDQSHLQETSSLRKLLDVENLCKTNSNCDVTRQEL
AKSKKQORVSESNTVDESNTNWVDGQSLSNSDDDEKASVTSVKGKTRATKGTATDPQSL
YARKRREKINERLKTQLNLPNGTKVDISTMLEAVHYVKFLQLQIKLLSSDDLWMYAPL
AYNGLDMGFHNNLLSRML*

>G23 (22..732)

TATCAAACGAGAGTACAAAAGATGACGCTCACTCAACAGCTCTGCATCACCAACATCATCG
 TCATCAGACCAATCTGATGCACTACTACAACAAGCACCCACTTGTCTGAAGAAGAAGCT
 CCACCCAGAAACAACAACAAGAAAGAGAAGGAGAGATTCTTCTTCTGCTTCTTCATCT
 TCTTCAATGCAACATCCTGTTTACAGAGGTGTGCGGATGAGAAGTTGGGGCAAATGGGTC
 TCCGAGATCCGACAACCTCGTAAGAAAACCTCGTATTTGGCTCGGCACCTTTTGTACCGCT
 GATATGGCTGCTCGTCTCACGACGTCGCTGCTCTCACCATCAAAGGCTCCTCCGCCGCTC
 TTAAATTTCCCTGAGCTTGCTTCTCTCTTCCCTCGTCCGGCGTCATCATCGCCGCATGAT
 ATCCAGACAGCCGCCGAGAAGCCGCCCGCCATGGTGGTGAAGAAAACTGTTAGAGAAG
 GATGAGGCTCCGGAGGCCACCTTCGTGGAATCTTCTTACGTGGCGGCGGAGTCAGAG
 GATGAGGAGAGGTTTGAGAAAATTGTGGAGCTGCCTAACATTGAAGAAGGAAGTTATGAC
 GAGAGTGTGACATCACGTGCTGATCTGGCTTATTCTGAGCCGTTGATTGTTGGGTGTAT
 CCTCCGGTTATGGATTTTTATGAAGAAATATCGGAGTTTAAATTCGTGGAATTGTGGAGC
 TTTAATCACTAATTAAGTTAGGAAAGTGCATTATATGCAATATGTCATCATAGATAACA
 TTGTATTCTTTTCTTTTGTACGGATACGTAGCATATGCTACTATACTAGGGCTAGTG
 TACCAAATATTGTAAATATACCTTATTAATATTTATGTAAATGTGTAATATATATAACAT
 ACAATTATTGAAGTTTGAAATTGGAACTATCGTTACGCAATGTTCTTGTAAAAAAA
 AAAAAA

>G23 Amino Acid Sequence (domain in AA coordinates: 61-117)
 MTSLSNSASPTSSSSDQSDATTTTSTHLSEEEAPPRNNNRKRNRDSSSSSSSMQHPV
 YRGVRRMSWGKVVSEIRQPRKKTRIWLGTFTADMAARAHDAALTIKSSAVLNFPELA
 SLFPRPASSPHDIQTAAEAAMVVEKLLLEKDEAPEAPSSSESSVAAESEDEERLEK
 IVELPNIEEGSYDESVTSRADLAYSEPPDCWVYPPVMDFYEEISEFNFVELWSFNH*

>G2313 (104..724)
 CGTCGACACAATCGCTCTCCGTAACATATTCCACAAAACGATCTTCTTGTTCCTTGAAT
 TTTTAGCCATCTCTTTTTTTTTTTTCTCATTCTCGGATACTATGGCTTCGAGTCCACG
 CTGGACGGAGGACGACAACAGGCGTTTTAAGTCAGCTCTGTGCAATTCCCTCCGGATAA
 CAAGCGTTTGGTGAATGTGCGCCAGCATCTGCCGAAACCTTTGGAGGAGGTGAAGTACTA
 CTACGAAAAGTTGGTCAACGATGTTTATCTGCCGAAACCTTTAGAGAATGTCACCCAGCA
 TCTGCAGAAACCTATGGAATGGAGGAGATGAAGTACATGTACGAAAAGATGGCCAACGA
 TGTTAATCAGATGCCCGAGTACGTACCACTGGCGGAATCGAGTCAGTCCAAACGCAGGAA
 GAAGGATACGCCAAATCCTTGGACAGAGAAGGAACACAGATTGTTTCTGCAAGGATTGAA
 AAAGTATGGGGAAGGAGCTTCGACGTTGACATCAACGAATTTTGTGAAGACAAAGACTCC
 ACGGCAAGTGTCAAGCCATGCACAGTATTACAAAAGGCAAAAATCGGACAATAAGAAGGA
 GAAACGCCGAGTATTTTTGACATAACTTTGGAGTCTACCGAGGGCAATCCAGATTCTGG
 AAATCAGAACCTCCGGATGATGATGATCCGTCCCAAGGTCAAGGCATTGTCTTGGAGT
 TTAGATGTTGGAAGATAGAAGAATGGTGTGAAAGC

>G2313 Amino Acid Sequence (domain in AA coordinates: TBD)
 MASSPRWTEDDNRRFKSALSQPPDNKRLVNVAQHLPKPLEEVKYYYEKLVDVYLPKPL
 ENVTOHLQKPMEMEEMKMYEKMANDVNQMPEYVPLAESSQSKRRKKDTPNPWTEEBHRL
 FLQGLKKYGEASTLTSTNFVKTKTPROVSSHAQYKQKSDNKKERRSIFDITLESTE
 GNPDSGNQNPDDDDPSQGQGTCLGV*

>G2344 (1..573)
 ATGACTTCTTCAATCCATGAGCTTTCTGATAACATTGGAAGTCATGAGAAGCAAGAACAG
 AGAGATTCTCATTTCACCAACCAATCCCTTCTGCAAGAAATTATGAATCAATTGTTACA
 AGTTTAGTCTACTACAGACCCGGGACTACAAATTCATGGCACCTGGACAATATCCATAT
 CCAGATCCTTACTACAGAAGCATATTTGCACCGCCTCCACAACCGTATACCGGGGTACAT
 CTACAGTTGATGGGAGTGCAGCAACAAGGCGTTCTTTTACCATCTGATGCAGTCGAGGAA
 CCTGTTTTTGTAAAGCAAGCAATACCACGATATACTAAGGCGCAGACAATCAAGAGCA
 AGACTTGAGTCTCAGAATAAAGTCATCAAGTCACGTAAGCCGTATTTGTCATGAATCTCGG
 CATTTGCATGCGATAAGACGACCAAGAGGATGTGGCGGGCGGTTTCTAAATGCCAAGAAG
 GAGGATGAGCATCACGAAGACAGTAGTCATGAAGAAAAATCCAACCTTAGCGCTGGTAA
 TCCGCCATGGCTGCTTCTAGTGGTACATCTTGA

>G2344 Amino Acid Sequence (domain in AA coordinates: TBD)
 MTSSIHESLSDNIGSHEKQEQRDSHFQPPIPSARNYESIVTSLVYSDPGTTNSMAPGQYPY
 PDPYYRSIFAPPPQPYTGVLHQLMGVQQQGVPLPSDAVEEPVFNQKQYHGLRRRQSR
 RLESQNKVIKSRKPYLHESRHLHAIRRPGRGGRFLNAKKEDEHHEDSSHEEKSNLSAGK
 SAMAASSGTS*

TCCTCACTCTCTCTTTTTCTCTAACCATAAAATCTCTTTGATCTCTTTCTCTGTGTTT
TGATAATGGAAAATGTTGGTGTTGGGATGCCGTTTTACGATTTAGGGCAAACAAGGGTTT
ACCACTCTTGTCTGATTTCACGATTATCGGCGGAGAGGTATCCGGTAGGGTTCATGG
ATTTACTGGGTGTTTCATCGTCATACACCACCATACGCCGTTGATGCATTTTCCGACCA
CACTTAATCTCGTCTCGAGCGAAGCTGTGAATGGAGATGACGAAGAAGAAGAAGATGGAG
AAGAACAGCAGCATAGACAAGAAGCGGTTTAAATTCACTAAATGAGTAGAAAGCAG
CGAAGAAGAAGGTGCCAAAAGTGTCACTACGAGGAGTGAGGTTCTTCATCTAGATG
ATGGTTATAAGTGGAGAAAAATACGGTCAAACACTGGTCAAAGACGCCCTTTTCCAAGAA
ATTATTACCGTTGCACAACAACTTGGTGTGACGTGAAGAAGAGAGTAGAGATCATTC

GTGATCCAAGCAGTGTAAATCACCACCTTACGAAGGTCAACATACTCATCCTCGTCCACTAC
TCATCATGCCCCAAGAAGGCAGCTCTCCATCCAATGGCTCAGCTTCTAGGGCCACATTG
GCCTCCCTACACTCCCTCCTCAGCTTTTAGATTACAACAACCAACAACAACAAGCGCCGT
CTTCTTTTGGAAACCGAGTACATTAACAGGCAAGAAAAAGGAATTAATCATGATGATGATG
ACGATCATGTTGTGAAGAAGAGTCGAACTCGGGATCTGCTGGATGGAGCTGGTTTAGTCA
AAGATCATGGCCTTCTTCAGGATGTTGTTCCCTCTCATATCATTAAAGGAAGAGTATTAGT
TAATCGCATAATTATGTAGCTAGCTAGCTAG

>G2517 Amino Acid Sequence (domain in AA coordinates: TBD)

MENVGVGMPFYDLGQTRVYPLLSDFHDLAERYPVGFMDLLGVHRHTPTHTPLMHFPTTP
NSSSSEAVNGDDEEEEDGEEQHKTKKRFKFTKMSRKQTKKKVPKVSFTRSEVLHLLDDG
YKWRKYGQKPVKDSFPFRNYRCTTTWCDVKRVERSFSDPSSVITTYEQHHTHPRPLLI
MPKEGSSPSNGSASRAHIGLPTLPPQLLDYNNQQQAPSSFGTEYINRQEKGINHDDDDD
HVVKSRTRDLLDGAGLVKDHGLLDVVP SHIIKEEY*

>G2521 (103..768)

ATTCTCCACAATTTTCATAACTTTCTTCCGCTCAACTTCAGATAAATTCGGATTCTGTAGC
TCTTTCAATACGACTGCGGAGATCAGAGCCAATTATTTGGTTATGGCGTCTCTGATCTCA
GATATTGAACCGCCGACGAGTACTACTTCAGATCTCGTTCGGAGAAAGAAGAGATCCTCT
GCTTCATCGCCCGCATCGTCTCGTTCAAGCGCATCTTCCGTCTCCGGTGAGATTACGCG
CGATGGCGATCGGAGAAGCAACAACGGATCTACTCAGCCAACTGTTCCAAGCGCTCCAA
CAAGTCCGCCTCAACTCTTCCGCCTCAACATCATCATCTCCAACGGCTCAGAAACGAGGA
AAGGCCGTCCGTGAAGCCGCGATCGAGCTCTTGCCGTTTCCGCTCGGGAAGAACAACCTC
TGGAGCAGAGCGATCTTAGCTAATCGGATCAAACTGAAATTTTCGTAAACAGAGACGTCCT
CGAGCTACGATGGCGATTCCGGCCATGACTACGGTGGTTAGTAGCAGCAGCAACAGATCG
AGAAAACGGAGAGTGTCTGGTGTGAGATTGAATAAGAAGAGTATACCGATGTTAACC
AAAGTACGTGTTCTAGGCCGTTAGTTCCCGGTTGCGGTAAACAATCCGTACCGGTGATT
CTAGAAGAAGCAACTGATTATATTAGGCTCTGGAGATGCAAGTGAGAGCCATGAATCT
TTAGTTACGCTTCTCTCTCTACGGCTCAGCTCCTCCACCGATTTGATGAGGTTAAAT
CGTCTTTTTTAATCTACCATCTCTCGATCTTTCACAGCTTATGTGTATATAGAAGATT
GTTTGATTATAATCTGTAATACTCTTCCCAACCGCTGATTCTTCTCTGCTACAAGTAAA
AGTAAATTTGAACCGAGTCTTCCATTTTACGATCCTCAAGTCTAAATTAAGTATATG
ATTGATTAATAAAGTCTTTACCATTAGGGTTC

>G2521 Amino Acid Sequence (domain in AA coordinates: 145-213)

MASLISDIEPTSTTSDLVRRKKRSSASSASSRVSSEIHRWRSEKQRIYSK
LFQALQOQVRLNSSASTSSPTAQKRGKAVREAADRALAVSARGRTLWSRAILANRIKLK
RKQRRPRATMAIPAMTTVVSSSNRSRKRVSVLRLNKKSIIPDVNRKVRVLGRLVPGCGK
QSVFVILEEATDIYIQALEMQVRAMNSLVQLLSYGSAPPI*

>G258 (60..983)

AGTGACCACCTGTGGTTAATCAACACCAAGAGACCTTGTAATATATAAGTTAGGAAGA
TGAGAGAGAAGTGGGAAATGAAAAGAGATGAAATGGGACATCGATGTTGTGGAAAACACA
AAGTGAAGAGAGGTCTTTGGTCTCCAGAGGAAGACGAGAAGCTTCTCGTTATATACCA
CTCATGGTCATCCTAGTTGGAGTTCGTTCCAAAGCTTGCCGGGTTGCAGAGATGTGGGA
AGAGTTGCAGATTAAGGTGGATAAACTATCTAAGGCCTGATCTGAGGAGAGGTTCTGTTA
ATGAGGAAGAAGAGCAGATTATCATCGACGTACATCGTATCTTGGTAACAAATGGGCTC
AGATTGCTAAGCACTTACCTGGACGCACTGATAATGAAGTCAAGAACTTTTGGAACTCAT
GCATTAAGAAGAACTTCTTTCTCAAGGCTTAGATCCTTCTACACATAATCTTATGCCTT
CACACAAAAGATCTTCTTCTCAAAACAATAATAATATCCCCAAGCCAAACAAAACGACGT
CCATCATGAAGAACCCTACTGATCTTGATCAATCAACCACTGCTTTTTCAATCACAACA
TCAATCCACCCACTTCCACTAAACCAAACTTAAATCTCCTAACCAGACTACAATCC
CATCTCAAACCGTGATCCCTATCAATGATAACATGTCAAGTACTCAAACCATGATCCCTA
TCAATGATCCCATGTCAAGTCTTTAGATGATGAGAATATGATTCCCTCACTGGTCAGATG
TTGATGGAATGGCGATCCACGAAGCTCCGATGTTGCCTAGTGATAAGGCAGTAGTGGGAG
TGGATGATGATGATCTCAACATGGACATTTTGTAAACTCCTTCTTCTCTGCTTTTG
ATCCTGATTTTGCTTCCATTTTCTCCTCTGCAATGTCTATCGATTTCAATCCCATGGATG
ATCTTGGCAGCTGGACCTTTTAGCTTTTACTCTACAGC

>G258 Amino Acid Sequence (domain in AA coordinates: 24-124)

MREKWEKMRDEMGRCCGKHVKRGLWSPEEDEKLLRYITTHGHPSSVPLAGLQRCG
KSCRLRWINYLRPDLRRGSFNEEEEQIIIDVHRILGNKWAQIAKHLPGRTDNEVKNFWNS

CIKKLLSQGLDPSTHNLMPSHKRSSSSNNNNIPKPNKTTSIMKNPTDLDQSTTAFSITN
INPPTSTKPNKLKSPNQTTIPSTQVIPINDNMSSTQTMIPINDPMSSLLDDENMIPHWSD
VDGMAIHEAPMLPSDKAVVGVDLNDILFNTSPSSAFDPDFASIFSSAMSIDFNPM
DLGSWTF*

>G280 (108..722)

AAGTTAATATGAGAATAATGAGAAAACCACTTTCCCAAATTGCTTTTTTAAATCCCTCCT
CACACAGATTCCCTTCCTTCATCACCTCACACACTCTCTACGCTTGACATGGCCTTCGATC
TCCACCATGGCTCAGCTTCAGATACGCATTATCAGAAGTCCGTCGTTTTCTCTCCAC
CTTATCTCAGATGATAATGGAAGCGATTGAGTCCTTGAACGATAAGAACGGCTGCAACA
AAACGACGATTGCTAAGCACATCGAGTCGACTCAACAACTCTACCGCCGTCACACATGA
CGCTGCTCAGCTACCATCTCAACCAGATGAAGAAAACCGGTCAGCTAATCATGGTGAAGA
ACAATTATATGAAACAGATCCAGATGCTCCTCCTAAGCGTGGTCGTGGCCGTCCTCCGA
AGCAGAAGACTCAGGCCGAATCTGACGCCGCTGCTGCTGCTGTTGCTGCCACCGTCTG
TCTCTACAGATCCGCTTAGATCTCGTGGCCGTCACCGAAGCCGAAAGATCCATCGGAGC
CTCCCCAGGAGAAGGTCATTACCGGATCTGGAAGGCCACGAGGACGACCCGAAGAGAC
CGAGAACAGATTCCGAGACGGTTGCTGCGCCGGAACCGGCAGCTCAGGCGACAGGTGAGC
GTAGGGGACGTGGGAGACCTCCGAAGGTGAAGCCGACGGTGGTTGCTCCGGTTGGGTGCT
GAATTAATCGGTACTTATGCAATTCGGAATCTTTAGTTACTGAAAAATGGAATCTCTTA
GAGAGTAAGAGAGTGCTTTAATTTAGCTTAATTTAGATTTATTTGGATTTCTTTTCAGTATT
TGGATTGTAACTTTAGAAATTTGTGTGTGTGTGTGTGCTTAGTCCTGAGATAAGATATAA
CATTAGCGACTGTGTATTATTATTACTGCATTGTGTATGTGAACTTTGTTCTCTT
GTTGAAAAA

>G280 Amino Acid Sequence (domain in AA coordinates: 97-104,130-137-155-162,185-192)

MAFDLHHGSASDTHSSELPSPFSLPPYPQMIMEAIESLNDKNGCNKTTIAKHIESTQQTLF
PSHMTLLSYHLNQMKKTGQLIMVKNYMKPDAPPKRGRRPPKQKQAESDAAAAAVV
AATVVSTDPSPRSRGRPPKPKDPSEPPQEKVITGSGRPRGRPPKRPRTDSETVAAPPEPAAQ
ATGERRRGRPPKPKPTVPAVGC*

>G3 (16..477)

GTTTGTCTTTTATCAATGGAAAGAGAACAAGAAGAGTCTACGATGAGAAAGAGAAGGCAG
CCACCTCAAGAAGAAGTGCCTAACCACGTGGCTACAAGGAAGCCGTACAGAGGGATACGG
AGGAGGAAGTGGGCAAGTGGGTGGCTGAGATTCTGTGAGCCTAACAAACGCTCACGGCTT
TGGCTTGGCTCTTACACAACCGATATCGCCGCCGCTAGAGCCTACGACGTGGCCGCTCTC
TACCTCCGTGGCCCTCCGCACGTCTCAACTTCCCTGATCTTCTCTTGCAAGAAGAGGAC
CATCTCTCAGCCGCCACCACCGCTGACATGCCCGCAGCTCTTATAAGGGAAAAAGCGGCG
GAGGTGCGCGCCAGAGTCGACGCTCTTCTAGCTTCTGCGGCTCCTTCGATGGCTCACTCC
ACTCCGCCGGAATAAAACCCGACTTGAATCAAATACCCGAATCCGGAGATATATAGTCA
ATTTATATACATGTAGTTTGTGTTTGTGTTGATTAGAAGATTACATTTACATACAAGATACA
CATAGATACTGGAAATATAGGTATGTATACATTCATAAATTATCTTATGTATCAAAGAA
TTTTATAGATTCTGATTAGCTTTTGTGTTTGTGTTTGTGATAAGAACTCTGATTAGTTGTC
CGGAGACAAAACCGGCTAAGAGCAATCCATGAGAAGCTAGCGAGTGTTTTTAGTTCAAG
TTGTAATATAAATGCATATTAATTCTTTAGTAATTTGT

>G3 Amino Acid Sequence (domain in AA coordinates: 28-95)

MEREQEESTMRKRRQPPQEEVPHVATRPYRGIRRRKWGWAEIREPNKRSRLWLGSY
TTDIAAARAYDVAVFYLRGPSARLNFDDLQEEHLSAATTADMPAALIREKAAEVGAR
VDALLASAAPSMASHTPPVIKPDNLQIPESGDI*

>G343 (1..795)

ATGGACGCTCTATGGETTATCTTCACAGACTTACTTCGAATCGACGACCTTCTTGATTTCT
TCCAACGAAGACATCTTCTCCGCTTCTTCTCCGGTGGTTCCACCGCCGCTACTTCTCTCT
TCTTCTTTCCCTCCTCTCAAACCCCTAGTTTCCACCAACCACCATCTCCCTTCTCTCCGCC
GATCATCACTCCTTCTCCACGACATTTGCGTTCCAGTGATGACGCAGCTCATCTTGAA
TGGCTTTTCGAATTCTGTGACGATTCTTTCGCTGATTTTCCGGCGAATCCATTAGGAGGA
ACTATGACTTCTGTCAAACTGAAACTTCTTTCCGGGGAAACCAAGAAGCAAACGATCA
AGAGCTCCTGCTCCTTTCCGCCGAACATGGTCTCCGATGCCACTGGAATCCGAGCATCAG
CAGCTTCACTCCGCCGCCAAATTCAAGCCAAAGAAAGAAACAATCCGGCGGAGGAGGAGGA
GGAGGAGGAAGACATGCATCGTCATCGGAGACTACGGAAGGAGGAGGAATGAGGAGA
TGTACTCACTGTGCATCGGAGAAAACGCCACAGTGGAGGACAGGACCCTTGGACCTAAA

ACACTATGTAACGCTTGTGGAGTCCGGTTTAAATCCGGTAGACTTGTACCGGAATATAGA
CCGGCTTCGAGTCTACTTTTGTGTTTACTCAGCATTCAAACCTCTACCGGAAAGTGATG
GAGCTTCGACGGCAGAAAGAAGTTATGAGACAACCACAACAAGTTCAACTTCATCACCAC
CACCACCCGTTTTAG

>G343 Amino Acid Sequence (domain in AA coordinates: 178-214)
MDVYGLSSPDLLRIDLLDFSNEDIFSASSSGGTAATSSSFPPPNPSFHHHLPSSA
DHHSFLHDICVPSDDAAHLEWLSQFVDDSFADFPANPLGGTMTSVKTETSFPGKPRSKRS
RAPAPFAGTWSMPLESEHQQLHSAAKFKPKKEQSGGGGGGGGRHQSSSSETTEGGGMRR
CTHCASEKTPQWRTGPLGPKTLCNACGVRFKSGRLVPEYRPASSPTFVLTHQHSNSHRKVM
ELRRQKEVMRQPQQVQLHHHHHPF*

>G363 (1..780)
ATGAGACCAATATTAGACCTCGAAATTGAAGCTTCATCGGGCAGTAGTAGCAGCCAAGTG
GCCTCAAACCTGTCTCCGGTTGGGGAAGATTACAAACCAATCTCGCTGAATCTTAGCCTC
AGTTTCAACAACAACAACAATAATCTGGATCTTGAATCATCGTCTTTGACGCTGCCA
CTTTCGAGCAGCAGTGAAGTAGTAACCCGAGCAGCAGCAGCAACAACAACCATCTGTA
TCAAAGAGAGTCTTCTCTGTAACCTACTGCCAAAGGAAGTTCTATAGCTCTCAAGCGCTA
GGTGGTCACCAAAACGCTCACAACGTGAGAGAACACTCGCCAAACGCGCTATGCTATGG
GTCTTGCTGGGGTCTTCCCGGTAGAGGATCAAGTAGCAATTATGCGGCTGCTGCCACAG
CAGCCGCTCTCTGTGTTTCCCGCTTACCGGAAGCGGAAACGGGAACATGACATCGTTTCAAG
ACTTTGGGAATCCGGGCACATTCTCGGCGCAGCAGCTCAGCATGACAAGGCAGACACCA
GAAACACTTATTAGAAACATTGCCAGGTTCAACCAGGGGTATTTTCGGTAATTGTATACCT
TTTACGTGGAGACGACGAGGCGGAGATGCTCTGGCCGGGGAGTTTCCGGCAAGCTACG
AATGCGGTTGCGGTTGAAGCGGTAATGATAATTTAGGTGAAAGAAAAATGGATTTCTTG
GACGTCAAGCAAGCGATGGATATGGAAGTTCTCTTCCAGATCTAACCTTGAAGCTTTGA

>G363 Amino Acid Sequence (domain in AA coordinates: 87-108)
MRPILDLEIEASSGSSSSQVASNLSPVGEDYKPISLNLSLFSNNNNNNLDLESSSLTLP
LSSTSESSNPEQQQQQPSVSKRVFSCNYCQKQFYSSQALGQHNAHKRERTLAKRAMLW
VLLGSSPVEDQVAIMRLLPQQPLSCLPLHSGNGNMTSFRTLGIRAHSSAHDVSMTRQTP
ETLIRNIARFNQGYFGNCIPFYVEDDEAEMLWPGSFRQATNAVAVEAGNDNLGERKMDFL
DVKQAMDMESSLPDLTLKL*

>G370 (1..774)
ATGGACGAAACCAACGACGAAGAGAAACTCAGGATTTTCATGAACGTCAACGTTGAATCC
TTCTCTCAGCTTCTTTTCATCCGCCGTACTCTCCCAAAGAAAAAGCCGCATTTATTCGT
CTCTTCGGCCAAGAGCTCGTCGGTGATAACTCCGACAACCTTATCCGCAGAACCTTCTGAT
CATCAAACCACTACCAAGAAGATGAGAGCTCTGAGAATATCAAGGACAAAGACAAAGAA
AAAGATAAGGACAAAGACAAAGATAACAACAACAACAGGAGATTTCGAGTGTCACTACTGC
TTCAGAAACTTCCCAACTTCTCAAGCCCTAGGTGGACATCAAACGCTCACAACCGTGAA
CGTCAACACGCCAAACGCGGTTCCATGACATCATACCTTCATCATCATCAGCCTCATGAC
CCTCACCATCATCTACGGCTTCTTCAACAACCACCACCGTCACTATCCGCTCTTGGACG
ACGGAAGCTAGATCATACTACGGCGGAGGGGACATCAAACGCCGTCGTACTACTCAAGG
AATACTCTTGCTCTCTCTTCTTAACCCACCGACAATCAACGGAAGTCTTTTAGGTTTG
TGGCGTGACCGCTTCCACGTCAACAAATACTATTAAGGCGTTTACTCATCTTACCA
GCTTCAGCGTTTAGGTGCGATGAGCAAGAGACTAATAAGGAGCCTAATAACTGGCCGTAC
AGATTGATGAAACCAATGTGCAAGATCATGTGAGTCTCGATCTTCATCTCTGA

>G370 Amino Acid Sequence (domain in aa coordinates: 97-117)
MDETNGRRETHDFMNVNVSFSQLPFIRRTPPKEKAAIIRLFGQELVGDNSDNLSEPSD
HQTTTTKNDESSNIKDKDKDKDKDKDNNNNRRFECHYCFRNFTSQUALGQHNAHKRE
RQHAKRGSMTSYLHHHQPHDPHHIYGFLLNNHHHRHYPSTTEARSYGGGGHQTPSYYSR
NTLAPPSSNPPTINGSPLGLWRVPPSTSTNTIQGVYSSSPASAFRSHEQETNKEPNWPY
RLMKPNVQDHSVSLDLHL*

>G385 (37..2202)
TAGGGTTTGCTTTCAGTTTCCGGAGTATAAGAAAAGATGTTTCGAGCCAAATATGCTGCTT
GCGGCTATGAACAACGACGACGCAATAACCACTACAACACGAGACAAACAATAAT
GAAGGATTTCTTCGGGACGATGAATTCGACAGTCCGAATACTAAATCGGGAAGTGAGAAT
CAAGAAGGAGGATCAGGAAACGACCAAGATCTCTTCATCTTAACAAGAAGAAACGATAT
CATCGACACACCAACTTCAGATCCAGGAGATGGAAGCGTTCTTCAAAGAGTGTCTTCAC
CCAGATGACAAGCAAAGGAAACAGCTAAGCCGTGAATTGAATTTGGAACCTCTTCAGGTC

AAATTCTGGTTCCAAAACAAACGTACCCAAATGAAGAATCATCACGAGCGGCATGAGAAC
TCACATCTTCGGGCGGAGAACGAAAAGCTTCGAAACGACAACCTAAGATATCGAGAGGCT
CTTGCAAATGCTTCGTGTCTAATTGTGGTGGTCCAAACAGCTATCGGAGAAATGTCATTC
GACGAACACCAACTCCGTCTCGAAAATGCTCGATTAAAGGAAGAGATCGACCGTATATCC
GCAATCGCAGCTAAATACGTAGGCAAGCCAGTCTCAAACCTATCCACTTATGTCTCTCTCT
CCTCTTCCTCCACGTCCACTAGAATCGCCATGGGAAATATTGGAGGAGAAGCTTATGGA
AACATCCAAACGATCTCTTAAGTCCATCACTGCACCAACAGAATCTGACAAACCTGTCTC
ATCATCGACTTATCCGTGGCTGCAATGGAAGAGCTCATGAGGATGGTTCAAGTAGACGAG
CCTCTGTGGAAGAGTTTGGCTTTAGACGAAGAAGAATATGCAAGGACCTTTCCTAGAGGG
ATCGGACCTAGACCGGCTGGATATAGATCAGAAGCTTCGCGAGAAAGCGCGTTGTGATC
ATGAATCATGTTAATACGTGAGATTCTCATGGATGTGAATCAATGGTCGACGATTTTC
GCGGGGATGGTTTCTAGAGCAATGACATTAGCGGTTTATCGACAGGAGTTGCAGGAAAC
TATAATGAGAGCTCTTCAAGTGATGAGCGCAGAGTTTCAAGTTCATCTCCATTAGTCCCA
ACACGTGAAACCTATTTTCGCACGTTACTGTAAACAACAAGGAGATGGTTCTGGGCGGTT
GTCGATATTTCTGGTGGATAGTCTCCAACCAATCCCCCGGCTAGATGCAGGCGGCGAGCT
TCAGGATGTTTGATTCAAGAATTGCCAAATGGATATTCTAAGGTGACTTGGGTGGAGCAT
GTGGAAGTTGATGACAGAGGAGTTCATAACTTATACAAACACATGGTTAGTACTGGTCAT
GCCTTCGGTGCTAAACGCTGGGTAGCCATTCTTGACCGCCAAATGCGAGCGGTTAGCTAGT
GTCATGGCTACAAACATTTCTCTGGAGAAGTTGGCGTGATAACCAACCAAGAAGGGAGG
AGGAGTATGCTGAAATTTGGCAGAGCGGATGGTTATAAGCTTTTGTGAGAGTGAGTGCT
TCAACCGCTCACACGTGGACTACATTGTCCGGTACAGGAGCTGAAGATGTTAGAGTGATG
ACTAGGAAGAGTGTTGATGATCCAGGAAGGTCTCCTGGTATGTTCTTAGTGACGCCACT
TCTTTTGGATCCCTGTTCTCTCCAAAGCGAGTCTTTGACTTCCTCAGAGACGAGAATTCA
AGAAATGAGTGGGATATTCTGTCTAATGGAGAGTTGTGCAAGAAATGGCACATATTGCT
AACGGGAGGGATACCGGAACTGTGTTTCTCTTCTTCGGGTAAATAGTGCAAACTCTAGC
CAGAGCAATATGCTGATCCTACAAGAGAGCTGCATTGATCCTACAGCTTCCTTTGTGATC
TATGCTCCAGTCGATATTGTAGCTATGAACATAGTGCTTAATGGAGGTGATCCAGACTAT
GTGGCTCTGCTTCCATCAGGTTTGTCTATTCTTCTGATGGTAATGCCAATAGTGGAGCC
CCTGGAGGAGATGGAGGGTCTGCTTGTGCTTTTTCAGATTCTGGTTGACTCAGTT
CCTACGGCTAAGCTGTCTCTTGGCTCTGTTGCAACTGTCAATAATCTAATAGCTTGCACT
GTTGAGAGAATCAAAGCTTCAATGTCTTGTGAGACTGCTTGAAAACCATCCATTAGC
>G385 Amino Acid Sequence (domain in AA coordinates: 60-123)
MFEPMMLLAAMNNADSNHNNHEDNNNEGFLRDDEFDSPNTKSGSENQEGSGNDQDPL
HPNKKRYHRHTQLQIQEMEAFFKECPHPDDKQRKQLSRELNLEPLQVKFWFQNKRTQMK
NHHERHENSRLRAENELRNDNLRYREALANASCPNCGGPTAIGEMSFDHQLRLNARL
REEIDRISAIKAYVGPVSNYPLMSPPPLPPRPLELAMGNIGGEAYGNPNPDLKLSITA
PTESDKPVIIDLSVAAMEELMRMVQDEPLWKSLLALDEEYARTFPRGIGPRPAGYRSEA
SRESAVVIMNHVNIVEILMDVNQWSTIFAGMVSRAVTLAVLSTGVAGNYNGALQVMSAEF
QVPSPLVPTRETYFARYCKQGGDGSWAVVDISLDSLQPNPPARCRRRASGCLIQELPNGY
SKVTWVEHVEVDDRGVHNLKHMVSTGHAFGAKRWVAILDRQCERLASVMATNISSGEVG
VITNQEGRSMLKLAERMVISFCAGVSASTAHTWTTLSGTGAEDVRVMTRKSVDDPGRSP
GIVLSAATSFWIPVPPKRVDFLRDENSRLNEWDLNNGGVVQEMAHIANGRDGTGNCVSL
RVNSANSQSNNMLILQESCIDPTASFVIYAPVDIVAMNIVLNGGDPDYVALLPSGFAILP
DGNANSAGPGGSGSLTVAFAQILVDSVPTAKLSLGSVATVNNLIACCTVERIKASMSCT
A*
>G439 (128..967)
TATAAATCTTCGTTTCTACTTTTTTTTCTTCCATAATATAGTCAATTCGTTTTCTTAATT
AGGGCTTCTTCTCTTGTCTTCTCAATCTTTATTAGTTTATTATTATTGTTATTG
TATACAAATGGCAATGGCTTTAAACATGAATGCTTACGTAGACGAGTTCATGGAAGCTCT
TGAACCATTCATGAAGGTAACCTCATCTTCTTCTACTTCAATTCATCAAATCCAAAC
ATTAACCTCCTAATTCATCCCTAATAATGACCAAGTCTTACCGGTATCTAACCAACCGG
TCCGATTGGGCTAAACAGCTCACTCCAACACAAATCTTCAAAATTCAGACAGAGTTACA
TCTCCGGGCAAAACCAATCTCGTCTCGCTGGTAGTCATCTTCTCACCCTAAACCAAC
CTCAATGAAGAAAATCGACGTAGCAACTAAACCGTTAAACTATACCGAGGCGTAAGACA
GAGGCAATGGGGTAAATGGGTAGCTGAGATTCCGGCTACCTAAACCAACCGGTTATG
GCTCGGTACGTTTCAAGACGGCTCAAGAAGCTGCATTAGCTTACGATCAAGCAGCTCATAA
GATCAGAGGAGACAACGCTCGTCTCAATTTCCAGACATTGTTCTGTCAGGACACTATAA

ACAGATATTGTCTCCGTCTATCAACGCAAAGATCGAATCCATCTGCAATAGTTCTGATCT
TCCACTGCCTCAGATCGAGAAACAGAAACAAACAGAGGAGGTCTCTCTGGTTTTTCCAA
ACCGGAGAAAGAACCGGAATTTGGGGAGATATACGGATGCGGATACTCGGGCTCATCTCC
TGAGTCGGATATAACGTTGTTGGATTTCTCAAGCGACTGTGTGAAAGAAGATGAGAGTTT
CTTGATGGGTTTGCACAAGTATCCTTTCTTGGAGATTGATTGGGACGCTATAGAGAACT
CTTCTGAATCCATTTTATLCTTTTGTATTTCATTTGTCTCTAAATTGTAGAATTTTATTTTC
AGAGCTTTTGAAGGGAAGTTCTTGAATGAGAGTTGCAGAGGACTAGTGAACCTAACTCT
GTTTTCTTTTGAAGTATTGTTTATAATGGGCCGTTGAATGGGCCTTATTGATTTAAACA
GCCCAAGTTTTTAAAAAAAAAAAAAAAAAAAAAAAAA

>G439 Amino Acid Sequence (domain in AA coordinates: 110-177)
MAMALNMNAYVDEFMEALEPFMKVTSSTSSSNPKPLTPNFI PNNDQVLPVSNQTGPI
GLNQLTPTQILQIQTELHLRQNSRRRAGSHLLTAKPTSMKKIDVATKPVKLYRGVRQRQ
WGKWWAEIRLPKNRTRLWLTGFETAQEAALAYDQAAHKIRGDNARLNFPDIVRQGHYKQI
LSPSINAKIESICNSSDLPLPQIEKQNKTEEVLSGFSKPEKEPEFGEIYGCYSGSSPES
DITLLDFSSDCVKEDESFLMGLHKYPSLEIDWDAIEKLF*

>G440 (237..1301)

AAAAAATCAGTTTTCATAACACGTTTTTCTCTCTCACCCACCAAAAAAATCTTTTGT
TCTTGTACCAAAAAATCTCGTGATAAATCTCTCAAACTTTGTTTTATTTCTTCTTGA
TTCTCTCGAAATCTCTCTCAACAAACCCAGAACTTTCTTGATTTCGCAAGCTTTTCTTC
CTTTTATATTCTCTCATTTTGTGCGAATATAGAGAGAGTCCATAAAAGAAACAGTAATGG
ACGAATATATTGATTTCCGACCATTGAAGTACACAGAGCACAAGACTTCAATGACTAAAT
ACACCAAAAAGTCATCGGAAAACTTTCCGGTGGTAAGTCATTGAAAAAGGTTAGTATTT
GTTTACTGATCTGACGCAACAGATTCAATCAAGTGACGAAGACGAAGAAGATTTCTTGT
TTCTCGCCGGAGAGTCAAAAGATTCTGTTAACGAGATCACTGTTGAGCCTAGCTGTAACA
ACGTCGTCACCGGAGTTTCGATGAAAGATAGAAAGAGACTCTCTCTCTCCGATGAAA
CTCAATCTCCGGCGTCGAGTCGTCAACGTCCTAATAACAAAGTTTCAGTCTCCGGTCAGA
TAAAGAAGTTCCCGTGGTGTAGACAACCGCCATGGGGGAAATGGGCGGCGGAGATTAGAG
ATCCGGAGCAACGTCGGAGGATTTGGCTCGGGACTTTTGAACGGCGGAGGAAGCTGCCG
TGGTTTATGATAACGCCGCTATAAGACTCCGTGGACCGGACGCTTTAACTAATTTCTCCA
TACCGCCTCAAGAAGAGGAAGAAGAAGAAGAACCGGAACCGGTTATTGAGGAGAAACCGG
TTATTATGACGACGCCAACACCAACAACATCGAGTTCTGAATCAACTGAAGAAGATTTAC
AACATCTCTCATCTCTACTTCCGTTCTCAATCACCGGTCAGAAGAGATTCAACAAGTAC
AACAACCGTTTAAATCAGCTAAACCCGAACCGGGGTTTCAATGCACCATGTTGGCATA
CCGGGTTTAAATACCGGTTTGGTGAATCAGACGATTCATTTCTTGGTACTCCGTTTC
TTGACAATATTTCAATGAATCACCACCAGAGATGTCAATATTTGACCAACCAATGGATC
AAATTTTCTGTGAAAATGATGATATCTCAATGATATGTTGTTCTTGGGTGGTGAACTA
TGAACATTGAAGATGAGTTAACAAGTTCTAGTATCAAAGATATGGGTTCAACGTTTAGTG
ATTTTGTGATGATTCATTGATATCAGATCTATTAGTTGCTTAATATGATGATGAGAGTGAAG
AAGAAACCATCAAGCAAAATATCTATGGTGTGACTGAAAAATTTGGTGTTACTTTTTTTT
CTTTTCATAAGTTTCATGAGCTTTTTTGTCTTTTTTTTAAATAATTTATTTAGTTTGTCA
GGAGCTTGTAACAGTTTGGGAGAAATAGTGGAATAATAGTTTAAATTAATAAAAAAAAAA
AAAAAA

>G440 Amino Acid Sequence (domain in AA coordinates: 122-189)
MDEYIDFRPLKYTEHKTSMTKYTKSSEKLSGGKSLKKVSICTDPDATDSSSDEDEEDF
LFPRRRVKRFVNEITVEPSCNNVVTGVSMDRKLSSSSDETQSPASSRQRPNNKVSVSG
QIKKFRGVRQRPWGKWAABIRDPEQRRRIWLTGFETAEEAAVVDNAAIRLRGPDALTNF
SIPPQEEEEEEPEPVIIEKPVIMTPTPTSSSESTEEDLQHLSSPTSVLNRSEIIQQ
VQQPFKSAKPEPGVSNAPWWHTGFNTGLGESDSSFPLDTPFLDNFYFNESPPEMSIFDQPM
DQIFCENDDI FNDMLFLGGETMNI EDELTS SSIKDMGSTFSDFDDSLISDLLVA*

>G5 (417..1421)

TTTTTTTTTTCGAATCTCCCCCTAATCTGTTGTTTCTCGCTTCTTCTCTGTTAATCATC
TGCTTTTCAAAAAGAAAGAAAAAGAAAAATTCGATTTCTGGGTTTGTGTTTGTGCATACA
GAAAAAATCAAGCTTATGAATTTGTGTTTAAATTTTTTGTGTTTAAATTTGAAAGGCAGGTT
TTTTTCAGAACGAGATCGTTTTTCAAATTTCTCTGATTTTACCTCTTTTTTCTTCTTA
GATTTTAGTGAAATCGAGGGTGAAATTTTGTATCCCTCTTTTCGGATCTACACAGAGGTT
GCTTATTTTCAAACCTTTTAGATCCATTTTTTTTTTAAATTTCTCGGAAAAATCCCTGTTTC
TTTACTTTTTTATAAGTCTCAGGTTCAATTTTTTTCGATTCAAATTTTTATTTTAAATG

CAGCTGCTATGAATTTGTACACTTGTAGCAGATCGTTTCAAGACTCTGGTGGTGAACCTCA
TGGACGCGCTTGTACCTTTTATCAAAAGCGTTTCCGATTCTCCTTCTTCTTCTGCGAG
CGTCTGCGTCTGCGTTTCTTACCCCTCTGCGTTTCTCTCCCTCTCTCCCGGTTATT
ACCCGGATTCAACGTTCTTGACCCAACCGTTTTCATACGGGTCGGATCTTCAACAAACCG
GGTCATTAATCGGACTCAACAACCTCTCTTCTTCTCAGATCCACCAGATCCAGTCTCAGA
TCCATCATCTCTTCTCCGACGCATCACAACAACAACCTCTTCTCGAATCTTCTCA
GCCCCAAGCCGTTACTGATGAGCAATCTGGAGTCTGCTGGATCTTGTTCGCTTACGGTT
CAGGTGTTCTTCTGAAGCCGACGAAGCTTTACAGAGGTGTGAGGCAACGTCACCTGGGGAA
AATGGGTGGCTGAGATCCGTTTGGCGAGAAATCGGACTCGTCTCTGGCTTGGGACTTTTG
ACACGGCGGAGGAAGCTGCGTTGGCCCTATGATAAGGCGGCGTACAAGCTGCGCGGCGATT
TCGCCCCGGCTTAACCTTCCCTAACCTACGTACATAACGATTTCACATCGGAGGCGATTTCG
GTGAATATAAACCTCTTCACTCCTCAGTCGACGCTAAGCTTGAAGCTATTGTATAAAGCA
TGGCGGAGACTCAGAAACAGGACAAATCGACGAAATCATCGAAGAAACGTGAGAAGAAGG
TTTCGTCGCCAGATCTATCGGAGAAAGTGAAGGCGGAGGAGAATTCCGTTTCGATCCGTG
GATCTCCACCGGTGACGGAGTTTGAAGAGTCCACCGCTGGATCTTCCGCTTGTTCGGACT
TGACGTTCCGTGACCCGAGGAGCCGCCGAGTGAACGAGACGTTCTCGTTGGAGAAGT
ATCCGTCGTACGAGATCGATTGGGATTTCGATTCTAGCTTAGGGGCAAAATAGGAAATTC
GCCGCTTGCAATGGAGTTTGTGAAATTCGATGACTGGCCCAAGAGTAATTAATTAAT
ATGGATTAGTGTAAATTTTCGTATGTTAATATTTGTATTATGGTTTGTATTAGTCTCTCT
GTGTCCGTCAGCTTGGGTTTTCGAGGCTCGACCATGCCACAGTTTTCATTTTATG
TAATCTTTTCTTTTGTCTTATGTAATTTGTAGCTTCAGTTTCTTCATCTATAATGCA
ATTTTATTATGATTATGTG
>G5 Amino Acid Sequence (domain in AA coordinates: 149-216)
MAAAMNLYTCSRFSQDSSGELMDALVPFIKSVSDSPSSSSAASASAFHPSAFSLPPLPG
YYPDSTFLTQPFYSYSDLQQTGSLIGLNNLSSSQIHQIQSQIHPLPPTHNNNSFSNL
LSPKPLLMKQSGVAGSCFAYGSGVPSKPTKLYRGVRQRHWGKWWAEIRLPRNTRLWLGT
FDTAEEAALAYDKAAYKLRGDFARLNFNLRHNGFHIGGDFGEYKPLHSSVDKLEAICK
SMAETQKQDKSTKSSKKREKKVSSPDLSEKVKABENSVSIGGSPVTEFEESTAGSSPLS
DLTFADPEBPPQWNETFSLEKYPSEIDWDSILA*
>G550 (1..1374)
ATGGCTGATCCGGCGATTAAGCTCTTTGGAAGACGATTCCCTTACCTGAGCTTGGTGT
GTTGATTCTTCTTCTAGCTATACCGGATTTTAAACCGAAACTCAGATTCTGTTCCGTTA
TCAGATTCTGTGTACCGCGATGATGATGATGAAGAGATGGGTGATTCCGTTTAGGACGA
GAAGAAGGTGATGATGTTGGTGATGGTGAGGAGAGCGAGACTGATAAAAAGGAAGAA
AAAGATAGTGAGTGTCAGGAAGAGTCATTGAGGAATGAATCTAATGATGTTACTACTACT
ACATCGGGTATAACTGAAAAACCGAAACAACAAAGCTGCAAAGACGAATGAAGAGTCA
GGTGGTACTGCTTCTCTCAAGAGGGGAAGTTAAAGAAACCTGATAAGATTCTACCGTGT
CCGCGATGTAACAGCATGGAACCAAGTTCTGTTACTACAACAACTATAATGTTAACCAA
CCTCGCCATTTCTGCAAGAAATGTCAGAGATATTGGACAGCTGGTGGAACGATGAGGAAT
GTTCCGGTTGGTGCTGGGAGACGTAAGAATAAGAGTCCAGCTTCTCATTATAACCGTCAT
GTAAGTATAACATCTGCGGAAGCTATGCGAAGGTGGCGAGAACTGATCTTCAACATCCT
AATGGTGCAAACTCTTCTCACTTTTGGCTCTGATTCTGTGCTTTGTGAATCTATGGCTTCT
GGATTGAATCTTGTGAGAAGTCATTGTTGAAGACACAACTGTATTGCAAGAACCCAAT
GAAGGCTTGAAGATTACCGTTCCGTTAAACCAGACAAACGAAGAAGCTGGAACAGTCAGC
CCGTTACCAAAAGTTCCATGCTTTCCAGGACCACCACCAACTTGGCCTTACGCTTGAAC
GGAGTTTCTGTGGACGATTTTACCGTTTACCCTCCACCGGCTTACTGGAGCTGCCCGGG
GTTTACCGGGGGCATGGAACAGCTTACATGGATGCCACAACCAATTACCATCTGGT
TCCAATCCAAATCTCTACTAGGTAAACATTACGTGACGAGAACGCTGCTGAACCA
GGAACCGCTTTTGTGAAACCGAGTCACCTGGTAGGGAGAAAAGCAACCCGAGAGATGC
TTGTGGGTTCCCAAGACGCTGAGGATTGATGATCCAGAGGAAGCTGCTAAAAGTTCCATC
TGGGAAACATTAGGGATCAAAAAGACGAAAATGCGGATACTTTCCGAGCTTTTCAGATCA
TCAACCAAGAAAAAGCAGTCTTTCTGAAGGAAGACTTCCGGGAAGAAGACCGAGTTG
CAAGCGAATCCTGCTGCTTTCTAGGTCAGCAAACTTCCATGAGAGCTCATAG
>G550 Amino Acid Sequence (domain in AA coordinates: 134-180)
MADPAIKLFGKTIPLPELVVDSSSSYTGFLTETQIPVRLSDSGDDDEEMGDSGLGR
EEGDDVGGGGESETDKKEKDESECQESLRNESNDVTTTTSGITEKTETTKAAKTNEES
GGTACSQEGKLLKPKDILPCPRCNSMETKFCYNNYNVNQPRHFCCKKCQRYWTAGGTMRN

VPVAGARRKNKSPASHYNRHVSITSAEAMQKVARTDLQHPNGANLLTFGSDSVLCESMAS
GLNLVEKSLKKTQTVLQEPNEGLKITVPLNQTNEEAGTVSPLPKVPCFPGPPPTWPYAWN
GVSWTILFPYPPPAYWSCPGVSPGAWNSFTWMPQPNPSGSGNPNSPTLGKHSRDENAAEP
GTAFDETESLGREKSKPERCLWPKTLRIDDPBEAAKSSIWETLGIKKDENADTFGAFRS
STKEKSLSEGLRPGRRLOANPAALSRANSFHESS*

>G670 (28..1152)

CACAGCATTGCAGCTGTGAATAACTAAATGGGGAGACATTCTTGCTGTTACAAAACAAAAG
CTGAGGAAAGGGCTTTTGGTCTCCTGAAGAAGACGAGAAGCTTCTTACTCACATACCAAT
CACGGCCATGGCTGCTGGAGCTCTGTCCCTAAACTCGCTGGTTTCGAGAGATGTGGGAAG
AGTTGTGCGACTCGAGCAGATCTGGTACCGCCGACTAAGATGGATCAATTACTTTGAGACCT
GATTTAAAGAGAGGAGCTTTTTCTCCTGAAGAAGAGAATCTCATCGTCAAGCTTTCATGCT
GTCCTTTGGAAACAGATGGTCAAGATTGCGTCAAGGCTTCGGGTGATGAACCGACAACGAG
ATCAAGAATCTATGGAATCAAGCATCAAGAAGAACTGAAACAAGAGGCATTGACCCA
AACACACACAAGCCCATCTCTGAAGTGGAGAGTTTGTAGCGACAAAGACAAACCAACA
AGCAACAACAAAGAAGCGGTAACGATCACAAGTCTCCTAGTTCCCTCTTCTGCGACTAAC
CAAGACTTCTTCTCGAAAGGCCATCTGATTTATCCGACTACTTCGGATTTCAGAAGCTT
AACTTCAACTCCAATCTAGGACTCTCTGTTACAAC TGATTCTTCACTCTGCTCGATGATT
CCGCCGCGAGTTTAGCCCCGGGAACATGGTTGGTTCTGTCTCCTTCAGACACCAGTATGCGTA
AAGCCCTCGATTAGTCTTCTCTCCGACACAACAGTTCAGAGTCTATCTCCGGAGGAGAT
CATGTGAAATTTGGCTGCACAAACTGGGAATTTTCAGACAAACAACATAATACCTCAAAAT
TTCTTCGACAATGGCGGATTCTCATGGTCTATCCCAAATCTTCTTACTTCTTCTTCACAA
GTCAAACCAAATCATAACTTCGAAGAAATAAAATGGTCAGAGTATTTGAACACACCGGTTCT
TTCATAGGGAGTACTGTACAGAGTCAAACCTCTCAACCAATCTACATCAAATCAGAAACA
GATTACTTAGCCAATGTTTCAAAACATGACAGATCCTTGGAGCCAAACGAGAACCTTGGGC
ACAAC TGAAACTAGTGACGTGTTCTCCAAGGATCTTCAGAGAAATGGCCGTCTCTTTTGGT
CAGTCCCTTTAGCTTTTTTTCTTTCTTTCTTTCTTATTCTAACAGATGTAGAGAACATAA
AGATATACAAATACATACAATGTCAATCGTACAGTGAGTAAAGTGTTCTGTATATTC
ATGGGCGAGCTGTCTTTTATGTTTAAAAA

>G670 Amino Acid Sequence (domain in AA coordinates: 14-122)

MGRHSCCYKQKLKRLGWSPEEDEKLLTHITNHGHGCWSSVPKLAGLQRCGKSCRLEQIWY
RRLRWINYLRPDLKRGAFSPEEENLIVELHAVLGNRWSQIASRLPGRTDNEIKNLWNSSI
·KKKLKQRGIDPNTHKPISEVESFSDKDKPTTSNNKRSNDHKSPPSSSATNQDFFLERPS
DLSDYFGFQKLNFNNSNLGLSVTTDSSLCSMIPQFSPGNMVGSVLQTPVCVKPSISLPPD
NNSSSPISGGDHVKLAAPNWEFQTNNNNTSNFFDNGGFSWSIPNSSTSSSQVKPNHNFEE
IKWSEYLNTPFFIGSTVQSQTSQPIYIKSETDYLANVNMTPWSQENLGTETSDVFS
KDLORMAVSGOSL*

>G760 (175..1878)

TGCTTAATTCCAATGCCATCGTGATCGATTATCTCTCTCTCTCTTCCAATTTCCCA
 ATTCTTTTTTAAAAACCCTAATTTTTTCAGATATCTGATTATCTCTTGATTTCCTTCTACTC
 GATTTGCTCCCATAAAAACCCTTACTTTCTTCAAGTTCTGGTTTTCCCGATTGATGGGT
 CGTGGCTCAGTGACGTCGCTTGCTCCTGGGTTCCGTTTTCACCCGACGGATGAGGAAGT
 GTTCGCTACTACCTTAAGCGTAAGGTCGCAACAAACCCTTTAAGTTCGATGCTATTTCC
 GTCACCGACATATACAAGTCTGAGCCTTGGGATCTACCAGATAAGTCGAAGCTGAAAAGT
 AGAGACTTGGGAATGGTACTTCTTTAGTATGCTGGATAAGAAGTACAGTAATGGTTCCAAG
 ACGAATCGTGCTACGGAGAAAAGGTTATGGGAAGCAGCTGGGAAGAAGATCGGGAGATTCTG
 AATGGTTCAAGACTCGTTGGGATGAAGAAGACACTGTTTATCAACAGGCTCGAGCTCCT
 CGTGGTGAAGGACCAATTGGGTTATGCATGAGTATCGGCTTTCTGATGAGGACTTGAAG
 AAAGCTGGTGTGCCACAAGAAGCATATGTGTTATGTAGGATATTCCAGAAAAGTGGTACG
 GGTCTTAAGAATGGGGAGCAGTATGGTGCCTCTTATCTTGAGGAGGAGTGGGAAGAAGAT
 GGAATGACTTATGTACCTGCTCAAGATGCTTTCAAGTGAAGGATTGGCTTTGAATGATGAT
 GTTTATGTGATATTGATGACATTGACGAGAAGCCCGAAAATCTGGTGGTCTATGATGCC
 GTTCTATTCTACCTAACTATGTGTCATGGGGAATCAAGTAACAATGTTGAATCAGGCAAT
 TACTCAGACTCTGGAAATTACATTCAACCAGGAAACAATGTTGTGCGACTCTGGTGGGTAC
 TTTGAACAACCAATTGAAACTTTTGAGGAAGATCGGAAGCCTATTATACGGGAGGGTAGC
 ATTCAGCCTTGTCTCTGTTTCCAGAGGAACAATTTGGCTGTGGTGTCGACGACGAAAAAT
 GTGGTGAATCTGGAATCTTCCAACAATAATGTGTTGTAGCTGATATCTGCTACGATGAC
 ATTCCTATTGATGATAAATACTTTACCCGATGAGCCATTACGGATCCTAATAACAATCTT

CCACTCAACGATGGTCTGTACCTGGAAACGAATGATCTCAGCTGTGCTCAACAAGATGAT
TTTAACCTCGAAGATTATCTCAGCTTCTTTGATGATGAGGGTTGACTTTTGACGATTCT
CTATTAATGGGACCTGAAGATTTTCTTCCCAACCAAGAAGCCCTTGACCAGAAACCTGCC
CCTAAAGAATTGGAGAAGGAGGTGCGCAGGAGGCAAAGAGGCAGTGGAGGAAAAGGAAAGT
GGCGAAGGATCTTCTTCAAAACAAGATACAGATTTCAGGACTTTGATTACAGCTCCGAAG
TACCCATTCTCAAAAAGACGAGCCACATGCTTGGAGCCATTCTACTCCATCTTCATTT
GCTTCACAGTTCCAAACAAAGGACGCAATGCGTCTACACGCAGCACAATCTTCTGGTTCA
GTTACGCTGACTGACAGGTATGATGAGAATATCAAACATGACTCTAGCAGCGGACAGCGGT
ATGGGCTGGTCAATGACAAGAACGGTAACCTCAACGTAGTCCTTTTATTCTGGGGTAGTC
CAACAGGATGATGCGATGACTGCCTCGGGAAGCAAGACAGGAATTACGGCGACAAGAGCT
ATGTTAGTCTTTCATGTGTTTATGGGTTCTCTACTCTCTGTTAGCTTCAAAATAGTAACC
ATGGTGTCTGCTCGGTAATAGGATCAAAGTTGAATCGTCTCAAAGACTTTTCTGGTGT
TGTACCTCTCAATCATATAGCCTTTAACTTTGGCAGTGCCTTGTGCTCAATATTTAA
TTTTAAAAAAAAAAAAAAAAA

>G760 Amino Acid Sequence (domain in AA coordinates: 12-156)
MGRGSVTS LAPGFRFHPITDEELVRYLKRKVCNKPFFDAISVTDIYKSEPWDLPDKSKL
KSRDLEWYFFSMLDKKYSNGSKTNRATEKGYWKTGKDREIRNGSRVVMKKTLVYHKGR
APRGERTNWVMHEYRLSDEDLKKAGVPQEAAYVLCRIFQKSGTGPKNGEQYGAPYLEEEWE
EDGMTYVPAQDAFSEGLALNDDVYVDIDIDEKPENLVVYDAVPILPNYCHGESNNVES
GNYSDSGNYIQPNNNVVDSSGYFEQPIETFEEDRKPIIREGSIQPCSLFPFEEQIGCGVQD
ENVNLESSNNNFVADTCYSIDIPIDHNYLPDEFMPPNNPLPLNDGLYLETNDLSAQQ
DDFNFEYDLSFFDDEGLTFDDSLMGPEDFLPNQEAALDQKPAKPELEKEVAGGKEAVEEK
ESGEGSSSKQDTFKDFDSAPKYPFLKKTSHMLGAIPTPSSFASQFQTKDAMRLHAAQSS
GSVHVTAGMMRISNMTLAADSGMGWSYDKNGNLNVVLSFGVVQDDAMTASGSKGITAT
RAMLVFMCLWVLLLSVSFKIVTMVSAR*

>G831 (92..1987)

TTCTTTTCATCGTTGTGTCTATTATAAATATATGTCAATTGGTTTCTAAAAATTCTACC
ATTGATTGATTGATTTTTTTTCTTTTAAGAGATGAATTTATTTACAAGAATCTCATCTCG
GACTAAGAAGGCCAATCTTTACTACGTAACCCCTAGTTGCTCTTCTCTGCATCGCTAGCTA
CCTTCTCGGTATTTGGCAAAACACGGCGGTTAATCCACGCGCCGCTTCGATGATTGAGA
CGGTACACCGTGCGAGGGATTACACAGACCTAATTCTACGAAAGATCTCGACTTCGACGC
GCATCACAACATTCAAGATCCACCTCCGGTGACGGAACCGCCGTAGTTTCCCGTCGTG
TGCCGCGCGCTTGAGCGAGCACAGCCATGCGAAGACGCGAAGCGATCGTTGAAATTCTC
GAGGGAGAGATTGGAGTATAGGCAAAAGGCATTGTCCCGAGAGAGAAGAAATCTTGAAGTG
CAGAATTCCGCGCGCTACGGTTACAAAACGCGCTTCCGATGGCCGCGAGTCGTGACGT
GGCGTGGTTGCTAATGTGCTCACACGGAGCTTACGGTTGAGAAAAAGAATCAGAATTG
GGTCCGGTACGAGAATGATCGGTTTTGGTTCCCTGGTGGAGGTACGATGTTTCCACGTGG
CGCTGATGCTTACATTGATGATATCGGACGGTTGATTGATCTCAGCGACGGCTCTATCCG
TACAGCCATCGATACCGGTTGCGGGGTGGCTAGCTTCCGTTGCATATCTTTTATCAAGAAA
CATTACAACGATGTCATTTCACCAAGAGACACACGAAGCTCAAGTCCAGTTCGCACT
CGAGCGTGGTGTGCGGCGGATGATCGGAATCATGGCTACAATCCGCTTACCGTACCTTC
TAGAGCCTTTGATTTAGCACATTGCTCTCGTTGCTTATTCCGTGGGGCCAAAACGATGG
GGCTTACTTGATGGAGGTGGATAGGGTTTAAAGACCAGGAGGGTACTGGATACTTTCTGG
ACCGCCGATTAATTGGCAGAAACGGTGAAAGGGTGGGAACCGACCATGGATGATTTGAA
TGCAGAGCAGACTCAGATCGAGCAGGTGCGGAGAAGCTTGTGTTGGAAGAAAGTTGTTCA
AAGAGATGATCTTGCTATTGGCAAAAACCTTTAACCACATTGACTGTAAGAAAACAG
AGAGGTTTTGAAAAATCCGGAGTTTTGTCTCATGATCAAGATCCCGACATGGCCTGGTA
TACGAAGATGGATTGTTGTTGACACCATTACCTGAAGTTGATGACGCTGAGGATCTAAA
GACGGTGGCCGGAGGGAAGGTAGAAAAGTGGCCGGCTAGATTAAACGCGATTCTCCGAG
AGTAAACAAAGGCGCTCTCGAGGAAATCACACCTGAAGCTTTCTGGAGAACACGAACT
GTGGAACACAGAGAGTTTCTTATTACAAGAAGTTAGATTACAGTTGGGTGAAACGGGAG
ATACAGAACTTAGTCGACATGAACGCTTACCTCGGTGGATTTCGCGGCGGCTCTAGCGGA
TGATCCGGTCTGGGTGATGAACGTTGTCCCGTGGAGGCTAAGCTCAATACGCTCGGTGT
CATCTACGAGCGTGTCTAATCGGAACGTATCAAACCTGGTGTGAAGCCATGTCGACGTA
TCCAAGAACGTATGATTTTATCCATGCTGACTCGGTTTTTACATTGTACCAAGGTCAATG
TGAACCGGAGGAGATATTGTTGGAGATGGACCGAATTCTTAGACCGGGTGGTGGTGTGAT
TATAAGAGATGACGTGGACGTTTTGATCAAGGTTAAGGAATTAACCAAAGGATTAGAATG

GGAAGGTAGAATTGCTGACCACGAGAAGGGTCCTCATGAAAGAGAGAAGATTACTATGC
GGTGAAACAGTATTGGACCGTTCTGCGCCTGATGAAGATAAAAAACAACACTAGTGCTCT
CTCCTGATTTTTGAGTTTTTTTTTCTTACAATGTTTTTTTTTTTTTTTCAATTTTT
TATACAACAATAAATTCTCAATAATTGTTGTCGCGGCCG

>G831 Amino Acid Sequence (domain in AA coordinates: 470-591)
MNLFTRISSRTKKANLYVTLVALLCIAASYLLGIWQNTAVNPRAAFDDSDGTPCEGFTRP
NSTKDLDFDAHNIQDPPPVTTETAVSFPSCAAALSEHTPCEDAKRSLKFSRERLEYRQRH
CPEREELKCRIPAPYGYKTPFRWPASRDVAWFANVPHTELTVKKNQNWVRYENDRFWF
PGGGTTFPRGADAYIDDIGRLIDLSDGSIRTAIDTGCQVASFAYLLSRNITMSTFAPRD
THEAQVQFALERGVPMIGIMATIRLPYPSRAFDLAHCSRCLIPWQNDGAYLMEVDRVL
RPGGYWILSGPPINWQKRWKGWERTMDDLNAEQTLQEQVARSLCWKKVVRDDLAIWQKP
FNHIDCKTREVLKNPEFCRHDQDPMAWYTKMDSCLTPLPEVDDAEDLKTAVAGGKVEKW
PARLNAIPPRVNGALEBITPEAFLENTKLWKQVSYKKLDYQLGETGRYRNLVDMNAY
LGGFAAALADDPVWVMNVVPVEAKLNTLGVIYERGLIGTYQNWCEAMSTYPRTYDFIHAD
SVFTLYQGQCEPEEILLEMDRILRPGGVIIIRDDVDVLIKVKELTKGLEWEGRIADHEKG
PHEREKIYYAVKQYWTVPAPDEDKNNTSALS*

>G864 (503..1534)
TGCAAAAACATTTTCTGTCTCTCCTCTGCCCAAATTTTTTTCTTTCCAGGAATATTTCT
CTAGAAAAACCAAGCAAAGCTTTAACCCTTCTCTCCTCAAAAGTAGCATCTTCTCTT
TTTCTATTTCTCCTTTCTCTTCTTATCTCTCTCTCGTTGTGAACGATTCCTTAAGAAT
ATAACCAAAGCCCTTTTCTCCTTTCTCAACTTTCCGGGAAAAATCTTCACGCAGCAAG
GTTTCTCTCTCGGCTCTCGCAGTGTTTTTTCGGGCCTTTTGTTCTTTCTATAAAAAA
TTCCGCTCCTTTAAGAAACTTTTTCCACCTAGAGAAGAAGAAGAGTATCACTCTTGTG
TTCAAGTTTCTCTCTTTAATAAAAAATCCATCTTTATTCTTTGTCTTTCTCTTTTGC
TTCCCTAATCTCTATGTTATAAACACACAGAGAGAAACAAAGTCACAGTCTCGAGTCAA
AAACAGAGAAATACGAAAGAAAAATGGAAGCGGAGAAGAAATGTTTCTACCGAGAATCAA
ATTCACAGAGACAAAACCAACACGACAACAATCGTATCGGAGTTAAACCAACTCACCA
AACCAGGATTCCTCGTATCTCAGTCACTGACCCAGACGCTACTGATTCCTCCAGTGACGA
CGAAGAAGAAGAACATCAACGCTTTGTCTCTAAACGCCGCTCGTGTTAAGAAGTTTGTCAA
CGAAGTCTATCTCGATTCCGGTGCTGTTGTACTGGTAGTTGTGGTCAAATGGAGTCGAA
GAAGAGACAAAAGAGAGCGGTTAAATCGGAGTCTACTGTTCTCCGGTGTTCAGCGAC
GAGACTACGACGGGAGAGAGAAAGTTCCGAGGAGTGAGACAGCGTCCATGGGGAAATG
GGCGGCGGAGATAAGAGATCCGTTGAAACGTGTACGGCTCTGGTTAGGTACTTACAACAC
GGCGGAAGAAGCTGCTATGTTTACGATAACGCCGCTATTACGCTTCGTGGTCCCGACGC
TCTGACTAATTTCTCAGTCACTCCGACAACAGCGACGGAGAAGAAAGCCCAACCGTC
TCCGGTGAAGAAGAAGAAGAAAAACAACAAAGCAAAAAATCCGTTACTGCTTCTTC
CTCCATCAGCAGAAGCAGCAGCAACGATTGTCTCTGCTCTCCGGTGTCTGTTCTCCGATC
TCCTTTCCGCGTCGAGCAATTTCTCCGGCATTTCTTCATCACCAGTCGCGGCCGTTGTAGT
CAAGGAAGAGCCATCCATGACCAACGGTATCTGAACTTTCTCTGATTTCTCGGCGCCCTT
GTTCTCAGATGATGACGTGTTTCGATTTCGGGAGCTCAGTGGTCCCGACTATCTCGGCGG
CGATTTATTTGGGAAGATCTATTACGGCGGATATGTGTACGGATATGAATTCGGATT
CGATTTCCGATCCGGATTATCCAGCTGGCACATGGAGGACATTTTCAAGATATCGGGGA
TCTATTCCGGTCCGATCCTCTTTAGCTGTTAATAATATTTTAAATAAATAAATAGTTA
TACCGGCCGTTACTAAACGGAACCGGAGAAAGTTTGTATACCGGTGACATAAAATCTCG
GTTATGTTTCTGTAATCTTTTTTCTTTGTTATATATAAAAAATGAATGAACTGAATTAA
TGTAAGTTAATGTTGATAATTATTACGTTTAAAGTTTGAIAAAAAAAAAAAAAAAAAA
AAAAAA

>G864 Amino Acid Sequence (domain in AA coordinates: 119-186)
MEAEKMMVLPRIKFTEHKINTTIVSELINTHQTRILRISVTDPDATDSSSDDEEEHQR
FVSKRRRVKFKFVNEVYLDGAVVTGSCGQMESKKRQKRAVKSESTVSPVVSATTTTTGK
KFRGVRQRPWGKWAAEIRDPLKRVRLWLGTYNTEEAAMVYDNAAIQLRGPDALTNFSVT
PTTATEKKAPPPSPVKKKKKKNNKSKSVTASSSISRSSNDCLCSRVSVLRSPPFAVDEF
SGISSSPVAADVVEEPSMTTVSETFSDFSAPLFSDDVDVDFRSSVVPDYLGGDLFGEDL
FTADMCTDMNFGPFDGSLSSWHMEDHFQDIGDLFGSDPLAV*

>G884 (31..1575)
TTTTTTTTTGTGTTAATTTTGGGGATCGATGTCGGAAAGGAAGAAGCTCCGTCGACA
TCGAAGTCCACCGAGCTCCGTCGCGTCCGACTTTATCTCTTCTCCACGGCCGTTTAGT

GAGATGTTCTTTAACGGTGGCGTTGGATTGAGTCTGGTCCGATGACTCTGGTCTCTAAT
ATGTTCCCTGATTCCGATGAGTTTAGGTCTTTCTCTCAGCTTCTCGCTGGAGCCATGTCT
TCTCCAGCGACTGCAGCTGCTGCTGCTGCTGCTGCGACGGCTAGTGATTACCAGAGACTT
GGTGAAGGGACTAATAGCTCTAGTGGTGATGTTGACCCGAGATTCAAGCAAAACAGACCA
ACCGGTTTGATGATTTCTCAATCTCAATCGCCGTCGATGTTACCGTACCGCCTGGTTTA
AGTCCAGCTATGTTGCTCGATTACCAAGCTTTTGGGTCTTTCTCTCCCGTTCAGGGA
TCATATGGAATGACACATCAGCAAGCTCTAGCTCAAGTCACTGCTCAAGCAGTTCAAGCC
AATGCCAATATGCAACCACAAACAGAGTACCCTCCTCCCTCTCAAGTTCAATCATTTTCA
TCGGGTCAAGCGCAGATCCCGACCTCGGCTCCACTACCAGCTCAAAGAGAAACCTCAGAT
GTAACCATCATAGAGCACAGGTACACACAGCCTCTAAATGTTGACAAACCAGCTGATGAT
GGCTATAACTGGCGAAAATATGGGCAAAAGCAAGTTAAAGGTAGCGAGTTTCCACGAAGC
TATTACAAGTGTACTAATCCAGGATGTCTGTCAAGAAGAAGGTTGAGAGATCTCTTGAT
GGACAAGTAACGGAGATTATCTACAAAGGTACAGACAATCATGAACCTCCTCAAAACACT
AAGCGAGGTAACAAAGATAACACCGCAATATAAATGGGAGTTCCGATAAATAACAATCGC
GGGAGTTCTGTAATTGGGGGCATCAGTTTCAAATAAGTCTCAACAAGACTAAGAGA
GAGCAACATGAAGCAGTAAGTCAAGCTACGACAAACAGAGCACTGTCTGAGGCAAGTGAC
GGTGAAGAAGTTGGTAATGGAGAACTGATGTGAGAGAGAAAGATGAGAATGAGCCTGAT
CCCAAGAGAAGAAGTACAGAAGTTCCGATTTTCAAGACCAGCTCCTGCTGCTTCACATAGA
ACTGTGACAGAGCCTAGAATTATTGTCCAAACGACGAGTGAAGTTGATCTTCTAGATGAT
GGATATAGGTGGCGTAAATATGGACAGAAAGTTGTCAAAGGGAATCCTTATCCGAGGAGC
TACTACAAGTGCACAACACCGAGATGTGGTGTGAGGAAACATGTAGAGAGAGCAGCAACA
GATCCAAAAGCTGTAGTAACAACATATGAAGGAAAACATAACCATGACCTTCCCGCTGCT
AAATCAAGCAGCCATGCCGCTGCAGCGGCACAGTTAAGGCCAGATAATCGACCTGGCGGT
TTGGCTAACTTAAATCAACAGCAGCAGCAACAGCCGTTGCGCGCTAAGGCTTAAAGAA
GAGCAACAACCTTGAGAGAAGAAACTCTTGACCGTTTTTTCATTACAAAGCTTTCAAAT
TCCACTCACACACTTGTCTGAAAAATCTAGCAGTTTGCAGGAAAGAAACAGCTTCAAGAG
GTTGTAGTTCTTCTATGTTCTGGTGTAAACTTAAAGCTTTTGGGGTTTTCAGATTTCT
TGTTTACTAATACTGTATGTGAATCTTTTGTACATGAGGAAGAAAATTACAGGGGGATA
TTTTGTGTGTATCTTTGTGTTATTGTTTCAGTAAAGATAGGTCTTACATTTTGTGTA
AAAAAAAAAAAAAAAAAAAA

>G884 Amino Acid Sequence (conserved domain in AA coordinates:227-285, 407-465)

MSEKEAPSTSKSTGAPSRPTLSLPPRFSEMFNNGVGFSPGPMTLVSNMFPDSDEFRS
FSQLLAGAMSSPATAAAAAAAAAATASDYQRLGEGTNSSSGVDPRFKQNRPTGLMISQSQS
PSMFTVPPGLSPAMLLDSPSFLGLFSPVQGSYGMTHQQALAQVTAQAVQANANMQPQTEY
PPPSQVQSFSSGQAIPTSAPLPAQRETSVDVTEIIEHRSQQPLNVDKPADDGYNWRKYGQK
QVKGSEFPRSYYKCTNPGCPVKKVERSLEDGQVTEIIYKGQHNHEPPQNTKRGKNDNTAN
INGSSINNNRGSSELGASQFQTNSSNKTREQHEAVSQATTEHLSEASDGEEVNGGETD
VREKDENEPPDKRRSTEVRRISEPAPAAASHRTVTEPRIIVQTTSEVDLLDDGYRWRKYGQK
VVKGNPYPRSYYKCTTPGCGVRKHVERAATDPKAVVTTYEGKHNDLPAKSSSHAAAAA
QLRPDNRPGGLANLNQQQQQPPVARLRLKEEQTT*

>G898 (161..772)

GAAGAAAAGATTCAAAAACCCTAGATTTCAAAAATCGATTGGCTGTCAAATTTCTCTCC
GGCGATTTTCTCGAGTGAAATTCGGCTCAAGGTGATTATAGCGATCATCGAATCAAAT
GATTGAAGAGGTACAAAGGTTAGTTACTTTGAGCTGAAAGATGAACACGTCAGAGGTGAG
AGTACCTCGAGGAAATCGACGGAGGAAAGCTGTGATTGATCTGAATGCGGTACCTGTTGA
TCAAGAAGGGACCTCTGCTTCTGTAGAACTCTTACGGTGCCTATTACACCGTCTCAGCC
TGCTCTACGATGATTGATGTCGATGCTATTGAGGATGATGTTATTGAATCATCCGCTAG
TGCTTTTGCTGAAGCTAAAAGCAAATCAAGAAATGCACGTCGGAGACCTTTGATGGTTGA
TGTAAGTGCAGGAGGTACGACTAGATTCCTGCCAACATAAGCAACAAACGAGAAGGAT
TCCTTCTAGTGAATCTGTGATCGACTGTGAGCATGCCTCTGTAAATGATGAAGTCAACAT
GTCTTCGAGAGTGTCTAGATCAAAGGCTCCAGCTCCTCCACCAGAAGAGCCAAAGTTTAC
ATGTCCAATCTGCATGTGTCCCTTTACGGAGGAGATGTCAACCAAGTGGGTACATCTT
CTGCAAGGGATGTATAAAGATGSCAATATCTCGCCAGGGCAAATGCCCTACTTGTAGGAA
AAAGGTTACTGCAAAAGAGCTGATTGAGTTTCTTCCAAACCACTAGATGAGTGGTCCG
GCAACATCACCAGCCACCCTGTCTAATGGTTTATCAGACTATCCTCCTATTCACTTTGGA
ACATTGAAGGGACTTCGTTGACTTGGTATTTTGAATATTTTGCTTTGTGGAAGAGAAA
TATTCAGTGATCAAGAAGCCAGAAGGCCCTATCATTGATGGATATCATTGGTAATAACT

CTTTGTTTTAGTTGTTGTTCTATGTAAATTTAGGTCTCTGCAAACCTCTCAGTCGATACT
CTTCTCTCTTGATAGATGATAAGATATATGGAAAAAAAAATTAATATTGAATCTTTACTA
AAA

>G898 Amino Acid Sequence (domain in AA coordinates: 148-185)
MNTSEVRVPRGNRRRKAVIDLNAVVDQEGTSASVRTLTVPITPSQPAPTMIDVDAIEDD
VIESSASAFEAESKSRNARRRPLMVDVESGGTTRFPANISNKRRIIPSSSEVIDCEHAS
VNDEVNMSSRVRSKAPAPPEEPKFTCPICMCPFTEEMSTKCGHIFCKGCIKMAISRQG
KCPTCRKKVTAKELIRVFLPTTR*

>G900 (1..648)

ATGGGGAAGAAGAAGTGCAGATTATGTTGTTGGTGTAGCGAGAATGTATTGTGAGTCAGAT
CAAGCGAGTTTATGTTGGGATTGTGACGGTAAAGTTCACGGAGCTAATTTTCTGGTGGCG
AAACACATGCGTTGTCTTCTATGTAGCGCGTGTGAGTCACACACGCCCTGGAAAGCTTCT
GGGCTGAATCTTGGCCCAACTGTTTCTATCTGTGAGTCTTGTTTAGCTCGTAAGAAGAAT
AACACAGCTCCCTCGCCGGGAGGGATCAGAATCTTAACCAAGAAGAAGAGATCATTGGT
TGTAACGACGGAGCTGAGTCTTATGATGAGGAAAGCGATGAGGATGAAGAAGAAGAAGAA
GTGGAGAATCAGGTTGTTCCGGCTGCGGTGGAGCAAGAAGTTCGGTGGTGAAGTTCGTCG
TCTTCGGTTAGTAGTGGTGAAGGAGATCAGGTGGTGAAGGACGAGACTTGATTTGGAT
CTTAACCTCTCCGATGAGGAGAACCAATCTAGACCATTGAAAAGATTATCGAGAGACGAA
GGTTTGTCAAGATCAACTGTTGTGATGAATAGCTCAATCGTGAAATTACACGGAGGGAGG
AGAAAAGCAGAGGGATGTGATACATCATCGTCTCTTCTGTTTATTGA

>G900 Amino Acid Sequence (domain in AA coordinates: 6-28, 48-74)
MGKKKCELCGVARMYCESDQASLCWDCDGKVHGANFLVAKHMRCLCSACQSHTPWKAS
GLNLGPTVSIKESCLARKKNSSLAGRDQNLNQEEIIGCNDGAESYDEESDEDEEEEEE
VENQVPAAVEQELPVVSSSSSVSSGEGDQVVKRRLDLNLSDEENQSRPLKRLSRDE
GLSRSTVVMNSSIVKLHGRRKAEGCDTSSSSSFY*

>G913 (108..806)

CATTCAAAAACATCATATATATACACAAACACACTTTGATACAACAAAAAACAGAAC
ACAAACAAAAACACATTGTAACATTAGTTTAAGCATTAAGCTTCTTTATGTGCAATAATA
ATAATTCTCCGACCACCGTGAATCAAGAAACGACGCTCTCGTGAAGTCTCAATCACAT
TGCTTACTGATCAATCTCCTCAAACTCACCAGGATCATCTTCTTCTCCTTACCCGAGAC
CTTCCGGTGGATCACCGGCGAGAAGAACGGCGACTGGATTATCCGGCAAGCACTCTATTT
TCAGGGGGATTGACTACGTAACGGAATGGGTATCGGAGATTAGAGAGCCACGTAATA
CGACAAGAATTTGGCTCGGGACTTATCCGGTACCGGAGATGGCTGCCGCCGCTTACGACG
TGGCTGCGTTAGCTTTAAAGGACCCGACGCCGTTTTGAATTTTCTGTTTAGCTTTGA
CTTACGTGGCTCCGGTTTCAAACCTCTGCTGCGGATATAAGAGCGGCTGCTAGTAGAGCAG
CGGAGATGAAGCAACCGGATCAGGGTGGGGATGAGAAGGTATTGGAACCGGTTCAACCCG
GCAAAGAGGAAGAATTAGAAGAAAGTGTCTGTAACTCGTGTTCGTTGGAGTTTATGGATG
AGGAAGCGATGTTGAATATGCCGACTTTGTTGACGGAGATGGCTGAAGGGATGTTGATGA
GTCCACCGAGAATGATGATACATCCGACGATGGAAGATGATTGCCCGGAGAATCATGAAG
GAGATAATCTTTGGAGTTATAAATGAATCCATTGAAGCTGCTCTTTTTTTATTGTTTTTC
CGGTGCAATGAGATTTTCCCCCTTTTTTTTTTTCTTTTTGGGTGCGTGT

>G913 Amino Acid Sequence (domain in AA coordinates: 62-128)
MSNNNSPTTVNQETTTREVSITLPTDQSPQTSPPSSSPSPRPSGGSPARRTATGLSG
KHSIFRGI RLNRGKWVSEIREPRKTTRIWLGTYPVPEMAAAAYDVAALALKGPDAVLNFP
GLALTYVAPVNSAADIRAAASRAEMKQPDQGGDEKVLPEVQPGKEEELEEVSNCNSCL
EFMDBEAMLNMPILLTEMAEGMLMSPPRMMIHPTMEDDSPENHEGDNLSYK*

>G937 (45..1046)

TGGAAAAAGTTTGAGTTTTTAATTCGAATCGAGAAAAATAAAATGGGTCTTTAGGTG
ATGAGCTTAGTTTGGGATCGATCTTTGGGAGAGGAGTTTCGATGAATGTTGTGGCGTTG
AGAAAGTTGATGAACATGTTAAGAAGCTTGAAGAAGAGAAGAGAAAGCTCGAAAGTTGTC
AACTTGAGCTTCTCTGTCTTTGTCAGATTTTAAACGATGCGATTTTGTATCTGAAGGATA
AGAGATGTTTCAGAGATGGAGACTCAACCATTGTTGAAAGATTTCAATTTCTGTTAATAAAC
CTATTCAAGGAGAAAGAGGAATAGAATTGCTGAAAAGAGAGGAGCTAATGAGGGAGAAGA
AGTTTCAGCAATGGAAGCTAATGATGATCAGTACTAGTAAGATCAAGAGCAAGCTTGAGA
TTAAGAGAAATGAGGAGAAATCTCTATGTTGTTGATTCCAAAGGTGGAACTGGTTTTAG
GCCTCGGTTTAAAGTTTCGAGTTTGATAAGAAGAAAAGGGATTGTTGCCTCATGTGGCTTTA
CTTCTAACTCTATGCCACAACCACCAACACCAGCAGTACCACAACAACCAGCATTCTTA

AGCAGCAAGCTTTACGGAAGCAAAGAAGGTGTTGGAATCCAGAGTTGCATCGCCGATTTG
TCGATGCATTGCAACAGCTAGGTGGACCGGGAGTGGCAACTCCTAAACAAATTAGAGAAC
ATATGCAAGAAGAAGGCTTAACCAATGATGAAGTCAAGAGTCATTTACAGAAATACAGGT
TACACATCAGGAAGCCAAATTCGAATGCGGAGAAACAATCAGCAGTTGTTTTAGGGTTTA
ACTTGTGGAATTCTTCAGCACAAGATGAAGAAGAGACATGTGAAGGAGGAGAATCATTGA
AGAGAAGCAATGCGCAATCAGATTCTCCTCAAGGTCCTTTGCAGTTACCGTCTACAACAA
CAACAACCTGGTGGAGATAGTAGCATGGAAGATGTTGAAGATGCTAAGTCTGAGAGCTTTC
AACTGGAGAGATTGAGATCACCATAAATCTCAAGAAACCAAACTCTTGATCACGGTTTTG
TTATTTTGGATTCTATTACTATATCTATTAGTAGTGAATGAGAACAATAATTATAGAAAGG
TTTATAGATATATATATAGAGAAAAAGAGAGAGTGGATGTTCAAATTATTTGCAGA
>G937 Amino Acid Sequence (conserved domain in AA coordinates:197-246)
MGS LGDELSLGSIFGRGVSMNVVAVEKVDEHVKKLEEEKRKLESCQLEPLSLQILNDAL
LYLKDKRCSEMETQPLLKDFISVKNPIQGERGIELLKREELMREKKFQQWKANDDHTSKI
KSKLEIKRNEEKSPMLLIPKVETGLGLSSSSIRRKGIVASCFTSNSMPQPPTPAVPQ
QPAFLKQQAALRKQRRWCNPELHRRFVDALQQLGGPGVATPKQIREHMQEGLTNDDEVKSH
LQKYRLHIRKPNSNAEQSAVVLGFNLWNSSAQDEEETCEGGESLKRNSNAQSDSPQGPLQ
LPSTTTTTGGDSSMEDVEDAKSESFQLERLRSP*
>G960 (63..1538)
TACCGTCGACCCACGCGTCCGAGTGATTCAAAGTCGGAAAGAAACCTAAAGAAGAGGA
TTATGGGTGCTGTATCGATGGAGTCGCTTCCTTTAGGTTTCAGATTACAGACCTACCGATG
AAGAGCTCCGTCAATCACTACCTCCGTCTCAAGATCAACGGACGTCACCTCCGATGTCCGTG
TCATCCCTGATATCGATGTCTGCAAATGGGAACCTTGGGATCTTCTGTCTCTCGGTGA
TTAAGACGGATGATCCAGAGTGGTTCTTTTTCTGCCCTCGTGATCGGAAATACCCTAATG
GTCATCGCTCTAACAGAGCAACTGACTCTGGCTATTGGAAAGCTACTGGTAAAGATCGTA
GCATCAAGTCTAAGAAGACTTTAATCGGTATGAAGAAGACTCTTGTCTTCTATCGTGGAC
GAGCTCCTAAAGGTGAGCGGACTAATTGGATTATGCACGAGTATCGTCCCCTCTTAAGG
ATCTTGATGGCACTTCCCCTGGCCAAAGCCCTTACGTTCTTTGTGCGCTCTTCCACAAGC
CTGATGATCGGGTTAATGGGTGTCAGTCCGATGAAGCAGCTTTTACGGCCAGCAACAAAT
ACTCACCTGATGATACATCATCTGATCTTGTTCAGAAACACCTTCTCTGATGCTGCTG
TTGAGAAACCATCAGATTATTCAGGTGGATGCGGTTATGCTCATAGTAATAGTACCGCAG
ATGGGACAATGATTGAGGCACCTGAAGAGAATCTTTGGTTATCTTGTGACCTTGAAGATC
AAAAGGCACCACTACCGTGTATGGATTCTATATATGCTGGTGATTTTCAATTACGATGAGA
TTGGATTCCAATTTCAAGATGGTACCAGCGAACCAGATGTATCACTAACAGAATTGTTGG
AGGAGGTGTTCAATAACCTGATGACTTCTCTTGGCAGGAATCGATCAGTCGAGAGAATC
CAGCAGTCTCACCAATGGGATATTTTCTCTGCTAAATGCTGCAGTCTGCAGCACCAG
AGGATGCTTTCTTCAACGACTTCATGGCTTTCAGTATACAGATGCTGAGATGGCGCAAT
TGCAGTATGGTTTCAAGGTGGAGCTTCTGGTTGGCCAAGTGACACTAATTCATACTATA
GTGATTTGGTTTCAAGAGCAAAATGATCAATCATAACACAGAGAACAACCTCACAGAAG
GGAGAGGGATAAAGATCCGGGCTCGACAGCCTCAGAACCAGCAGAGTACAGGATTGATAA
ACCAGGGTATTGCTCCAAGGAGAATCCGTCTGCAGCTGCAGTCTAATCTGAAGTAAAG
AACGAGAGGAGGTGAATGAAGGACACACTGTTATTCCCGAGGCCAAAGAAGCTGCAGCTA
AATACTCAGAGAAGAGTGGTTCTTTGGTTAAACCTCAAATAAAGCTCAGGGCGCGGGGAA
CTATAGGCCAAGTAAAGGAGAGAGATTGTCAGACGACGAGGTACAGGTGCAGAGCACAA
AGAGAGAGAGAGAGAGAATCAAATGTAGTTTAATGTAATTAGGGATGATGCAATGTTAGC
ATGTTTGTGTGTTGTAACCTTAAAACTTATTAGGAATCTGATAAAAGTTACTGTTGAAA
AAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
>G960 Amino Acid Sequence (domain in AA coordinates: 13-156)
MGA VSMESLPLGFRFRPTDEELVNHYLRLLKINGRHSDVRVIPDIDVCKWEPWDLPALSVI
KTDDPEWFFFCPRDRKYPNGHRSNRATDSGYWKATGKDRSISKKTLLIGMKTLLVFYRGR
APKGERTNWIMHEYRPTLKLDGTS PGQSPYVLCRLFKHPDDRNVGVKSDEAAPTASNKY
SPDDTSSDLVQETPSSDAAVEKPSDYSGGCGYAHNSNADGTMIEAPEENLWLSCDLEDQ
KAPLPCMDSIYAGDFS YDEIGFQFQDGTSEPDVSLTELEEVFNPNDDFSCBESISREN
AVSPNGIFSSAKMLQSAAPEDAFFNDFMAFTDTDAEMAQLQYGSSEGGASGWPSTNSYYS
DLVQQEQMINHNTENNLTGREGIKIRARQPQNRQSTGLINQGIAPRRIRLQLQSNSEVKE
REEVNEGHTVPIPEAKEAAKYSEKSGSLVKPQIKLRARGTIGQVKGERFADDEVQVQSTK
RERERIKCSLM*
>G991 (6..533)

GAAAAATGGAAGAAGAAAAGAGATTGGAGCTAAGGCTAGCTCCTCCTTGTACCAATTCA
CTTCCAACAACAACATCAATGGATCTAAACAAAAAGCTCGACCAAAGAAACATCATTCC
TTTCCAATAACAGGGTTGAGGTAGCTCCAGTGGTGGGATGGCCGCCGGTGAGATCATCCC
GGAGAAACCTAACGGCACAATAAAGGAGGAGATGAAGAAGAAGGAGAGTGATGAAGAGA
AGGAATTGTACGTTAAGATCAACATGGAAGGAGTCCAATAGGAAGAAAAGTCAACCTTT
CAGCTTATAACAACACTACCAACAGCTTTTACATGCCGTTGACCAACTCTTCTCTAAGAAAG
ATTCTGGGATCTAAACAGACAATACACTTTGGTCTACGAAGACACTGAAGGAGATAAAG
TTCTGGTGGGGATGTTCTTGGGAGATGTTTGTATCTACTGTAAAGAGGTTGCATGTTT
TAAAGACCTCCACGCCCTTCTCACTCTCACCTAGAAAAACATGGCAAGGAATAGAGAGAGG
TTGGCCAAAATCATCAGTTTCGATGGTTTGTCTTAAATGTAATTTTGTGGAACTAATGG
GGTTTGGCTTTGATTTACTGGTTTCTTTTCACTTATGTACTAGGTTTGTGCTTGCTAT
GTATTTCTTGTGTTTGGTTGTAATATGCTGTTTAAAGAAATCGGGGGTTAGTATGT
TATCGTGTGTATAAAAATAGTGTAAAGCACGTAAGTTGATTACAAAAAATAAAAAA
AAAAA

>G991 Amino Acid Sequence (domain in AA coordinates: 7-14,48-59,82-115,128-164)

MEEEKRLELRAPPCHQFTSNNNINGSKQKSSTKETSFLSNRVEVAPVVGWPPVRSRR
NLTAQLKEEMKKESDEEKELYVKINMEGVPIGRKVNLSAYNNYQQLSHAVDQLFSKKDS
WDLNRQYTLVYEDTEGDKVLVGDVPWEMFVSTVKRLHVLKTSFAFSLSPRKHGKE*

>G748 (98..1444)

CCACGCGTCCGCACTCTCCCAAATCTCTCTTCTTTAACAACAAAAAATCACAGAGA
CATAGAGAGAAGAAGACGGAACAGAGGCTCCAAAAAATGATGATGGAGACTAGAGATCC
AGCTATTAAGCTTTTCGGTATGAAATCCCTTTTCCGTCGGTTTTTGAATCGGCAGTTAC
GGTGGAGGATGACGAAGAAGATGACTGGAGCGGCGGAGATGACAAATCACCAGAGAAGGT
AACTCCAGAGTTATCAGATAAGAACAACAACACTGTAACGACAACAGTTTTAACAATTC
GAAACCCGAAACCTTGGACAAGAGGAAGCGACATCAACTGATCAGATAGAGAGTAGTGA
CAGCGCTGAGGATAATCAGCAGACGACACCTGATGGTAAACCCCTAAAGAAACCGACTAA
GATTCTACCGTGTCCGAGATGCAAAAGCATGGAGACCAAGTTCTGTTATTACAACAATA
CAACATAAACAGCCTCGTCATTTCTGCAAGGCTTGTCTAGAGATATTGGACTGCTGGAGG
GACTATGAGGAATGTTCTGTGGGGGAGGACGTCGTAAGAACAAGCTCATCTTCTCA
TTACCGTCACATCACTATTTCCGAGGCTCTTGAGGCTGCGAGGCTTGACCCGGGCTTACA
GGCAACACAAGGGTCTTGAGTTTGGTCTCGAAGCTCAGCAGCAGCAGTTGCTGCTCC
CATGACACCTGTTATGAAGCTACAAGAAGATCAAAGGCTCTAAACGGTGCTAGGAACAG
GTTTTCAGGGTTAGCGGATCAACGGCTTGTAGCTCGGGTAGAGAATGGAGATGATTGCTC
AAGCGGATCCTCTGTGACCACCTTAACAATCACTCAGTGGATGAATCAAGAGCACAAG
CGGCAGTGTGTTGAAGCACAATGAACAACAACAACAATAACATGAATGTTATGC
TTGCATCCAGGTGTTCCATGGCCTTACACGTGAATCCAGCGATGCCTCCACAGGTTT
TTACCCGCTCCAGGTATCCAATGCCGTTTTACCCCTTACTGGACCATCCCAATGCTACC
ACCGCATCAATCCTCATCGCTATAAGCCAAAAGTGTTCAAATACAACTCTCCGACTCT
CGGAAAGCATCCGAGAGATGAAGGATCATCGAAAAGGACAATGAGACAGAGCGAAAACA
GAAGGCCGGGTGCGTTCTGGTCCCGAAAACGTTGAGAATAGATGATCCTAACGAAGCAGC
AAAGAGTCCGATATGGACAACATTGGGAATCAAGAACGAGGCGATGTGCAAGCCGGTGG
TATGTTCAAAGGGTTTGATCATAAGACAAAGATGTATAACAACGACAAAGCTGAGAACTC
CCCTGTTCTTCTGCTAACCTGCTGCTCTATCAAGATCACACAATTTCCATGAACAGAT
TTAGAGTTACATATGTATATGTATATGTATGATTGATTGTATGTATAGATGATACTGG
AGAATGATGAGTTTTTGAGAATCAAACCTCTTTCTTTCTTAGTGATTGCCTTTATTCC
TTTACATGTTTGGTTCTCTGTACACTATTTGATTTACCTTTTTTACTTTCTTTCTTCAT
TTGTCAGGAAATGTTGGAAGATAACATTAATGGTAAAAAGTTGGTGTGGACCGTTGTTGC
GTTGGCATTTCAAAAAAATAAAAAA

>G748 Amino Acid Sequence (domain in AA coordinates: 112-140)

MMMETRDPAILFLGMKIPFVSFESAVTVEDEEDDWSGGDDKSPEKVTPELSDKNNNNC
NDNSFNNSKPELTLDKEEATSTDQIESSTPEDNQQTTPDGKTLKKPTKILPCPRCKSMET
KFCYYNNYINQPRHFCKACQRYWTAGGTMNRNVPVAGRRKNKSSSSHYRHITISEALEA
ARLDPLGLQANTRVLSFGLEAQQHVAAAPMTPVMKLQEDQKVSNGARNRFRHGLADQRLVAR
VENDGDCSSGSSSVTSSNNHSVDESRAQSGSVVEAQMNNNNNNNMNGYACIPGVWPYTNW
PAMPPPGFYPPGYPMPFYFYWTIPMLPPHQSSSPISQKCSNTNSPTLGKHPRDEGSSKK
DNETERKQKAGCVLPKTLRIDDPNEAAKSSIWTTLGIKNEAMCKAGGMFKGFDHKTMY
NNDKAENSPVLSANPAALSRSHNFHEQI*

>G247 (1..660)

ATGAGAATGACAAGAGATGGAAAAGAACATGAATACAAGAAAGGTTTATGGACAGTGGAA
GAAGACAAGATCCTCATGGATTATGTCCGAACCTCATGGCCAGGGCCACTGGAACCGCATC
GCCAAGAAAACCTGGGCTCAAGAGATGTGGGAAAAGCTGTAGGTTGAGATGGATGAACCTAC
TTAAGCCCCTAATGTTAACAGAGGCAATTTTACTGACCAAGAAGAAGATCTCATCATCAGA
CTCCACAAGCTCCTCGGCAACAGATGGTCGTTGATAGCGAAAAGAGTTCCGGGAAGAACA
GACAACCAAGTAAAGAATTACTGGAACACACATCTCAGCAAGAACTTGGTCTCGGAGAT
CATTCAACTGCCGTCAAAGCCGCATGCGGTGTAGAGTCTCCACCGTCTATGGCCCTTATA
ACCACAACGTCCTCCTCTCATCAAGAGATCTCCGGTGGAAAAAATTCAACTCTAAGGTTT
GACACTTTTAGTTGACGAATCCAACTCAAACCAAAATCCAACTAGTCCACGCAACACCA
ACTGACGTAGAAGTTGCAGCTACGGTTCCAAATCTGTTCGATACCTTTTGGGTTCTTGAA
GACGACTTCGAGCTTAGTTCACTCACTATGATGGATTTTACTAATGGGTATTGCTTTTGA
>G247 Amino Acid Sequence (domain in AA coordinates: 15-116)
MRMTRDGKEHEYKKGLWTVVEEDKILMDYVRTHGQGHWNRIAKKTGLKRCGKSCRLRWMNY
LSPNVNVRGNFTDQEEELIIRLHKLLGNRWSLIAKRVPGRDNDQVKNYWNTHLSKKLGLGD
HSTAVKAACGVESPPSMALITTTSSSHQEIISGGKNSTLRFDTLVDESKLKP KSLVHATP
TDVEVAATVPNLFDTFWVLEDDFELSSLTMMDFTNGYCL*

>G585 (111..2039)

CTCTCAAACATTTCTCTGTTTGTTCGGCGAAAACGGCAACTGTTTCATCAAATGACAAA
CACAAAACCTTAACATCTAGTTTGTATCCTCTCTGATACCTCAAAAAAATGGATGAAG
AAACAATGGCTACCGGACAAAACAGAACAACTGTGCCAGAGAATCTGAAGAAACACCTCG
CAGTTTCAGTTCGAAACATTCAATGGAGTTATGGTATCTTTTGGTCTGTCTCTGCTTCTC
AGTCTGGAGTTT TAGAATGGGGAGATGGATACTATAATGGAGATATCAAACGAGGAAGA
CGATTCAAGCTTCGGAGATCAAAGCTGATCAGCTTGGTCTACGGAGGAGCGAGCAGCTTA
GCGAGCTTTACGAGTCTCTCTCCGTCGCTGAATCTTCTTCTCAGGCGTTGCTGCCGGAT
CTCAAGTCACCAGACGAGCTTCCGCCGCCGCACCTTTCACCGGAAGATCTCGCCGACACCG
AGTGGTACTATTTGGTTTGTATGTTCTTTCGTTCTCAACATTGGTGAAGGAATGCCTGGAC
GGACGTTTGCAAACGGTGAACCGATATGGTTGTGCAACGCTCATACGGCGGATAGTAAAG
TGTTTTAGCCGTTCTCTCTAGCAAAAAGTGCTGCGGTTAAGACAGTGGTTTGCTTCCCGT
TCCTTGAGAGGAGTCGTTGAGATTGGTACCACAGAACATATTACGGAAGACATGAATGTAA
TACAATGCGTGAAGACATCATTCTCGAAGCCCTGATCCGTACGCTACAATATTACCAG
CAAGATCCGATTATCACATCGACAACGTTCTTGATCCGCAACAGATTCTAGGCGACGAGA
TTTACGCGCCTATGTTCACTACGGAGCCTTTTCCAACAGCTTCTCCGAGCAGAACTACCA
ACGGTTTCGATCAAGAACATGAACAAGTAGCAGATGATCATGATTCTTTTCATGACCGAAA
GAATCACTGGAGGAGCTTCTCAGGTGCAAAGCTGGCAGCTCATGGACGACGAGCTTAGTA
ACTGCGTTTACCAGTCTGCTAAATTCCAGCGATTGCGTCTCTCAAACGTTTGTGTAAGGGG
CGGCTGGACGGGTGCTTACGTTGCAAGAAAGAGTAGAGTTCAAAGACTAGGGCAAATTC
AAGAGCAACAGAGAAATGTGAAGACATTGTCAATTTGATCCAAGAAACGACGACGTTTCATT
ACCAAAGTGTGATCTCAACGATTTTAAAGACCAACCATCAGTTAATTCTCGACCGCAGT
TTCGAAACTCGGATAAACAGTCAAGCTTCACTAGGTGGAAGAAATCATCGTCATCATCAT
CAGGAACCGCCACGGTCACGGCACCATCAAGGAATGTTAAAGAAAATTATTTTCGATG
TTCCGCGAGTGCAACAGAAAGAGAAGTTAATGTTGGACTCACAGAAAGCCAGAGATGAAA
CTGGGAACCATGCGGTTTTAGAGAAGAAGCGCCGAGAAATTGAACGAACGGTTCATGA
CCTTGAGAAAAATCATTCCGTCAATCAACAAGATCGATAAAGTATCGATTCTTGACGATA
CGATAGAGTATCTTCAAGAACTCGAGAGACGGGTTCAAGAACTAGAACTTTCGAGAGAAT
CAACCGATACAGAGACTCGTGGGACGATGACGATGAAGAGGAAGAAACCATGCGACGACG
GAGAAAGAACATCAGCTAATTTGCGCAAATAATGAAACAGGAAATGGGAAGAAGGTGTCGG
TTAACAATGTTGGTGAAGCCGAGCCAGCAGATACCGGTTTACTGGTTTAACCGATAATT
TAAGGATCGGTTTCGTTTGGTAATGAGGTGGTTATTGAGCTTAGATGTGCTTGGAGAGAAG
GAGTATTGCTTGAGATAATGGATGTGATTAGTGATCTCCATTGGATTCTCATTCCGGTTC
AATCCTCGACCGGAGACGGTTTGTCTCTGCTTAACCGTCAATTGCAAGCACAAGGGGTCAA
AAATAGCGACACCAGGAATGATCAAAGAAGCACTTCAAAGGGTTGATGGATCTGTTGAA
GACTACTTAGTTAAATTTGACAGCAAAGAAAAACATTCCCGGTTTGGTTTCTATTCTTT
GGTTTTCTTCTAACCGGTTTTAGGAATTAATGTTATGTTTATCATTTGTTTTTGTGTTT
TTTTTTGTGTTCTTTTTTCCGTTGCTTAACGTAGGTGAAGAGGAACATACACTATGCGTA
TTTTGTTTGAGGTAGATTATTTAAGGGTATTAGTAATAGTAATAGCCAGTTTAGATGAT
TTTGTGTTCTTTTGTGTT

>G585 Amino Acid Sequence (domain in AA coordinates: 436-501)
MDEETMATGQNRRTVPENLKKHLAVSVRNIIQWSYGIFWSVSASQSGVLEWGDGYNGDIK
TRKTIQASEIKADQLGLRRSEQLSELYESLSVAESSSSGVAAGSQVTRRASAAALSPEDL
ADTEWYYLVCMSFVFNIGEGMPGRTFANGAPIWLCNAHTADSKVFSRSLAKSAAVKTVV
CFPFLGGVVEIGTTEHITEDMNVIQCVKTSFLEAPDPYATILPARSDYHIDNVLDPPQIL
GDEIYAPMFSTEPFPTASPSRTTNGFDQEHEQVADDHDSFMTERITGGASQVQSWQLMDD
ELSNVCVHQSLSNDDCVSQTVEGAAGRVAYGARKSRVQRLGQIQEQQRNVKTLSDPRND
DVHYQSVLSTIFKTNHQLILGPQFRNCDKQSSFTRWKKSSSSSSGTATVTAPSQGMKKI
IFDVPRVHQEKLMMLDSPEARDETCNHAVLEKKRREKLNRFMTLRKIIPSINKIDKVS
LDDTIEYLQELERRVQELSCRESTDTETRGTMKRRKKPCDAGERTSANCANNETGNGK
KVSNNVGEAEPADTGFTGLTDNLRIGSFGNEVIELRCAREGVLLEIMDVISDLHLDS
HSVQSSDGLLCLTVNCKHKGSKIATPGMIKEALQRVAWIC*

>G634 (1..798)
ATGGAGCAAGGAGGAGGTGGTGGTGGTAATGAAGTTGTGGAGGAAGCTTCACCTATTAGT
TCAAGACCTCCTGCTAACAACTTAGAAGAGCTTATGAGATTCTCAGCCGCCGCGGATGAC
GGTGGATTAGGAGGTGGAGGTGGAGGAGGAGGAGGAGGAAGTGCTTCTTCTCATCGGGA
AATCGATGGCCGAGAGAAGAACTTTAGCTCTTCTTCCGATCCGATCCGATATGGATTCT
ACTTTTCGTGATGCTACTCTCAAAGCTCCTCTTTGGGAACATGTTTCCAGGAAGCTATTG
GAGTTAGGTTACAAACGAAGTTCAAAGAAATGCAAAGAGAAATTCGAAAACGTTCAAGAA
TATTACAAACGTACTAAAGAACTCGCGGTGGTCTCATGATGGTAAAGCTTACAAGTTC
TTCTCTCAGCTTGAAGCTCTCAACTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
CATCAACCAGAACAGAAACAACAACAACCAACCAACAGAGATGGTCATGAGCTCGGAA
CAATCATCATTACCATCATCATCAAGATGGCCAAAGGCAGAGATTCTAGCGCTTATAAAC
CTGAGAAGTGAATGGAACCAAGGTACCAAGATAATGTACCTAAAGGACTTCTATGGGAA
GAGATCTCAACTTCAATGAAGAGAATGGGATACAACAGAAACGCTAAGAGATGTAAAGAG
AAATGGGAAAACATAAACAATACTACAAGAAAGTTAAAGAAAGCAACAACAGCAACTAC
AACAACAAGAATCAATGA

>G634 Amino Acid Sequence (domain in aa coordinates: 62-147, 189-245)
MEQGGGGGGNEVVEEASPISSRPANNLEELMRFSAAADDGGLGGGGGGGGGSASSSSG
NRWPREETLALLRIRSDMDSTFRDATLKAPLWEHVSRKLLLELGYKRSSKKCKEKFENVQK
YYKRTKETRGRHDGKAYKFFSQLEALNTTPPPPPSHPHAHQPEQKQQQPQEQEMVMSSE
QSSLPSSSRWPKAELALINLRSGMEPRYQDNVPKGLLWBEISTSMKRMGYNRNAKRCKE
KWEINKYKVKESNNSNYNNKNQ*

>G676 (1..612)
atgagaaagaaagtaagtagtagtggtgacgaaggaaacaatgagtacaagaaagggttg
tggaacagtagaagaagacaaaatcctcatggattatgtcaaagctcatggcaagggtcac
tggaatcgtattgccccaaagactggttttaagagatgtggaaagagttgtagattgagg
tggaatgaattatctcagccctaattgtgaaaagaggcaatttcaccgagcaagaaggat
cttatcattaggtccacaagttgcttggtaataggtggtctttaattgtcaaaagagt
ccgggtcgaacggataaatcaagtgagaactattggaacacgcacatcttagtaagaaactc
ggaaatcaaagatcagaaaacaaacagagcaatggtgatattgtttatcaaatcaatctc
ccgaatcctaccgaacatcagaagaaacgaaaatctcgaatattgtcgataacaataat
atcctcggagatgaaattcaagaagatcatcaaggaagtaactacttgagttcactttgg
gttcatgaggatgagtttgagcttagcacactcaccaacatgatggactttatagatgga
cactgtttttga

>G676 Amino Acid Sequence (domain in AA coordinates: 17-119)
MRKVVSSSGDEGNNEYKGLWTVEEDKILMDYVKAHGKGHWNRIAKKTGLKRCGKSCRLR
WMNYLSPNVKRGNFTEQEDLIIRLHKLLGNRWSLIAKRVPGRTDNQVKYWNTHLSKLL
GIKDKTKQSNQDIVYQINLPNTETSEETKISNIVDNNILGDEIQEDHQSNYLSLW
VHEDEFELSTLTNMMDFIDGHCF*

>G682 (1..228)
ATGGATAACCATCGCAGGACTAAGCAACCCAAGACCAACTCCATCGTTACTTCTTCTTCT
GAAGAAGTGAGTAGTCTTGTGAGTGGGAAGTTGTGAACATGAGTCAAGAAGAAGAAGATTG
GTCTCTCGAATGCATAAGCTTGTCTCGGTGACAGGTGGGAAGTATAGCTGGGAGGATCCCA
GGAAGAACCCTGGAGAAATTGAGAGGTTTTGGGTGATGAAAAATTGA

>G682 Amino Acid Sequence (domain in AA coordinates: 27-63)
MDNHRRTKQPKTNSIVTSSEEVSSLEWEVVMMSQEEDLVSRMHKLVGDRWELIAGRIP

GRTAGEIERFWVMKN*

>G635 (1..993)

ATGGAGATCATGCGTCCAGGGGTCTCAGAAAACACTTTGAAAGGAAAAATAAGAATCACA
ACGCGGTGCATGTGGCTTGACAAAGGAAGACTTTTAGATGCACCTTCAAAAGCAGCTCAT
GCTGCTCTATCAAGTTGTCTGTGACATGTCCCTGTCTCACATGGAAAGAACAGTCTCC
GAAGTCTTGAGGAAGATTGTAAGGAAGTACAGTGGTAAAAGGCTGAAGTCATCGCTATA
GCCACTGAGAATCCAAATGGCTGTCCGAGCTGATGAGGTCAAGTGCAGACTGTCTGGTGAT
CCAAGTGTGGTTCTGGAGTTGACGCTTTAAGGAAAGTTGTTGAAGGAAATGACAAAAGA
AGTCGGGCGAAGAAAGCACCTTCACAAGAAGCTTCCCCCAAAGTAGATCGCACTTTG
GAAGATGATATCATTGATAGTGCAAGACTACTGGCTGAAGAAGAACTGCGGCATCAACA
TACACGGAAGAAGTTGATACGCCCGTTGGGAGTTCTTCAGAAGAGTCAGACGATTTTTGG
AAATCATTTCATCAATCCATCATCGTCACCTTCACCGAGTGAAACAGAAAATATGAATAAG
GTAGCTGATACGGAGCCTAAAGCAGAGGGTAAGGAAAACAGCAGAGACGACGATGAATTA
GCTGATGCTTCAGATTCTGAAACCAAGTCATCACCAAAACGTGTGAGGAAGAAACAAATGG
AAACCGGAGGAGATAAAGAAGGTAAATCAGAATGCGAGGAGAGCTGCACAGTAGATTTCAA
GTGGTGAAAGGTAGAATGGCATTGTGGGAAGAGATCTCTTCAAATCTATCAGCTGAAGGA
ATCAATCGAAGCCCGGACAATGCAATCTCTCTGGGCATCACTTATTTCAGAAATACGAG
GAGAGCAAGGCTGATGAGAGAAGCAAGACGAGTTGGCCACATTTTGAGGATATGAACAAC
ATTTTGTCTAGAGCTAGGCACACCTGCGTCTTAA

>G635 Amino Acid Sequence (domain in AA coordinates: 239-323)
MEIMRPGVSENTLK GKIRITTRCMWLDKGRLLDALHKAHAALSSCPVTCPLSHMERTVS
EVLKIRKYSYSGKRPEVIAIATENPMVRADSVSARLSGDPVSGSGVAALRKVVEGNDKR
SRAKKAPSQEQASPKVDRTLEDDIIDSARLLAEETAASTYTEEVDTPVGSSEESDDFW
KSFINPSSSPSPSETENMNKVADTEPKAEGKENSRRDDELADSDSETKSSPKRVKRNKW
KPEEIKKVIRMRGELHSRFQVVKGRMALWEEISSNLSAEGINRSPGQCKSLWASLIQKYE
ESKADERSKTSWPHFEDMNNILSELGTPAS*

>G1068 (150..1310)

GAGAGTTGTTAGCTAGCTCACACGCTTTTCGCTTAAACTCAAAACCTGCACCTTTCTCGT
CTATTTTCTCGGCATTTCGTAAGAACAGAAAAGTGGGTCTCCAAGAAAATTACCCTAAATTC
ACAAAGATTCTACTTTTCTCCACCTCCAATGGATTCCAGAGAGATCCACCACCAACAAC
AGCAACAACAACAACAACAGCAGCAGCAGCAACAACAGCAACATCTACAACAACAGC
AAACACCACCGCCAGGGATGTTAATGAGTCACCACAATTCCTACAATCGAAACCCCTAACG
CCGCCGCCGCTGTTTAAATGGGTCAACAACCTCCACATCTCAAGCTATGCATCAAAGAT
TACCTTTTGGTGGTTCTATGTCACCGCATCAGCCTCAACAACATCAGTATCATCATCCTC
AGCCTCAGCAACAGATAGATCAGAAGACTCTTGAATCTCTTGGATTTCCTACTTCGCCTC
TTCCTTCTGCTTCTAATTTCTTACGGTGGTGGAAATGAAGGAGGTGGTGGTGGTAGCG
CCGAGCTAATGCTAATCTTCCGATCCACCTGCTAAACGGAACAGAGACGCTCCTCTG
GCTCCGGTAAGAAGCAGCTCGATGCTTTAGGAGGAACAGGAGGAGTTGGGTTACGCCTC
ATGTCATTGAGGTTAAACAGGAGAGGACATAGCTACGAAGATATTGGCGTTTACGAACC
AAGGCCACGCGCAATCTGTATTCTCTCAGCTACAGGAGCTGTAACATAATGTGATGCTTC
GTCAAGCTAACAATAGCAATCCTACTGGAAGTGAAGTATGAGGGCCGATTTGAAATCA
TTTCTCTGTCAAGTTCTTTCTTGAATTTCTGAGAGTAATGGTACTGTGACCAAACTGGTA
ACTTGAGTGTGTCGCTGGCTGGACACGAAGGCCGATTGTGGGTGGATGTGTGATGGAA
TGCTAGTAGCTGGATCAAGTCCAGGTCATTGTGGGAAGCTTTGTACCAGATGGAAGGA
AGCAGAAAACAAAGTCCGGGGCGTGCTCAGAATACTCCGGAGCCAGCTTCAGCACAGCCA
ATATGTTGAGCTTTGGTGGTGGTGGTGGACCGGAAGCCCTCGATCTCAAGGACAACAAC
ACTCGAGCGAGTCATCAGAGGAAACGAAAGTAATTTCTCCGTTCACCGTAGAAGCAACA
ACAACAACAGCAACAATCATGGGATATTTGGAACTCTACACCTCAACCGCTTCACCAAA
TTCCTATGCAGATGTACCAGAATCTCTGGCCTGGCAACAGTCTCAATAAACAGATGGTT
CATGGGTCAAGATTTGACCGGGTTTGCTTCTCTGTTCTTTTGACACATCTCTCCATCAG
ATTTATCTCTATAAAGTAGATTGAGCTCTCTTACTCTCTCATCTTCTTCTCTTTACTAT
TTCTCTTAAATTTAGCTTTGGTTTGTAGATAAATAGAGAGAGAGACATGTTAAGTAGGT
TTCAAATTCATCTTGTGTTAGTTGTTCTTAGTAGTTTCTTTGATTGTGATGATCATA
AAGACTTGTTCTTTTCTCTATATTCAACGAATTATCCACTTTAA

>G1068 Amino Acid Sequence (domain in AA coordinates: 143-150)
MDSREIHQQQQQQQQQQQQQQQHLQQQQQPPPGMLMSHHNSYNRNPNAAAVLMGHN
TSTSQAMHQRLPFGGSMSPHQPHQYHHPQPPQIDQKTLES LGFPTSPLPSASNSYGG

GNEGGGGDSAGANANSSDPPAKRNRGRPPGSGKKQLDALGGTGGVGFTPHVIEVKTGED
IATKILAFITNQGPRAICILSATGAVTNVMLRQANNNSNPTGTVKYEGRFEIISLSGSFLNS
ESNGTVTKTGNLSVSLAGHEGRIVGGCVDGMLVAGSQVQVIVGSFVPDGRKQKQSAGRAQ
NTPEPASAPANMLSFGGVGGPGSPRSQGGQHSSESSEENESNSPLHRRSNNNNSNNHGIF
GNSTPQPLHQIPMQMYQNLWPGNSPQ*
>G1225 (1..984)
ATGACTCTAGAAGCTTTATCATCAAACGGTCTTTTAAACTTTTGTCTCTGAAACTCTT
TCACCAACTCCATTCAGTCTCTCGTCGATCTCGAGCCATTGCCGGAATAATGATGTCTATC
ATATCGAAGAACAATTTTCGAGATATCTAATCAAGAACCACCACAGCGACAACCA
CCAGCTACGAATCGAGGGAAGAAGCGGCGGAGGAGGAAGCCTAGGGTTTGCAAAAACGAG
GAAGAAGCTGAGAATCAACGAATGACTCACATTGCCGTCGAAAGAAATCGAAGAAGACAA
ATGAATCAACATCTCTCTGTCTTGCATCTCTCATGCCTCAACCTTTTGTCTACAAGGGT
GATCAAGCTTCAATAGTTGGTGGAGCCATAGATTTTCATCAAAGAATTGAACACAAATTA
CTATCTCTTGAAGCTCAAAAACATCATAATGCTAAATTAAACAGTCGGTTACTTCTTCA
ACAAGTCAAGACTCAAATGGTGAACAAGAGAAATCCTCATCAACCATCTTCACTATCTCTA
TCGCAGTTCTTTCTTCAATTCATACGATCCGAGCCAAGAGAATAGGAACGGCTCAACAAGC
TCGGTGAAAACCCCTATGGAAGATCTTGAGGTGACTCTAATCGAACTCATGCTAACATC
AGAATCTTGTGCGAGAAGAAGAGTTTCCGGTGGAGCAGTTGGCCACCACCAACCGCCG
CAGCTTTCGAAGCTGGTGGCTTCTCTACAATCGTGTCCCTCTCCATTCTTCACTTAGT
GTCACAACTTGGACAATTATGCTATTTACTCCATCAGCGCTAAGGTGGAAGAGAGTTGC
CAGCTAAGTTCACTAGATGACATTGCAGGAGCAGTTCAACCATGCTAAGTATCATTGAA
GAGGAGCCTTTTGTGCTCATCAATGTCAGAATTACCATTGACTTCTCTTTGAATCAC
TCAAATGCTCACTATCTCTCTGAGAAATCTCTTTTGTGTTGTTATTCCTCTCTTTTA
ATTTTATCACATAGCACATCTTTAGTTTTTTTTTTT
>G1225 Amino Acid Sequence (domain in AA coordinates: 78-147)
MTLEALSSNGLLNFLLSETLSPTPFKSLVDLEPLPENDVIIISKNTISEISNQEPPPQRQP
PATNRGKKRRRRKPRVCKNEEEAENQRMTHIAVERNRRRQMNQHLVLRSLMPQPFHAKG
DQASIVGGAIDFIKELEHKLSSLEAQKHNAKLNQSVTSSTSQDSNGEQENPHQPSLSL
SQFFLHSDPSQENRNGSTSSVKTPMEDLEVTLIETHANIRILSRRRGFRWSTLATTKPP
QLSKLVASLQSLSLILHLSVTTLDNYAIYSISAKVEESCQLSSVDDIAGAVHMLSLIIE
EFPFCCSSMSELPDFSLNHSNVTHSL*
>G1337 (97..1398)
AATGGATTTGTTCATCATCTTCTCACCCTCCTTAGTCTCTGAAAATAAATCTGATTTTG
ATTTGCAATTTTAGGGATTTTGAGAGAGAGTCAGTTATGAGTAGTTCCGAGAGAGTACCG
TGCGATTTCTGCGCGAGCGTACGCGGTTTTGTTTTGTAGAGCCGATACGGCGAAGCTG
TGTTTGCCTTGTGATCAGCAAGTTACACGCGCAATCTGTTGTCGAGGAAGCACGTGCGA
TCTCAGATCTGCGATAATTCGCGTAACGAGCCAGTCTCTGTTCCGGTGTTCACCGATAAT
CTGATTTTGTGTCAGGAGTGTGATTGGGATGTTACGGAAGTTGTTCACTTTCCGATGCT
CATGTTGATCCGCGTGAAGGTTTTTCCGTTGTCCATCGGCGTTGGAGCTTGCTGCT
TTATGGGGACTTGATTTGGAGCAAGGAGGAAAGATGAAGAGAATCAAGTTCCGATGATG
GCGATGATGATGGATAATTTCCGGATGCAGTTGGATTCTTGGGTTTTGGGATCTAATGAA
TTGATTGTTCCAGCGATACGACGTTTAAAGAAGCGTGGATCTTGTGGATCTAGTTGTGGG
AGGTATAAGCAGGTATTGTGTAAGCAGCTTGAGGAGTTGCTTAAGAGTGGTGTGTCGGT
GGTGATGGCGATGATGGTGTGTCGACCGTGATTGTGACCGTGAGGGTGCTTGTGATGGA
GATGGAGATGGAGAAGCAGGAGAGGGGCTTATGGTTCCGAGATGTCAGAGAGATTGAAA
TGGTCAAGAGATGTTGAGGAGATCAATGGTGGCGGAGGAGGAGGAGTTAACCAGCAGTGG
AATGCTACTACTACTAATCCTAGTGGTGGCCAGAGTTCTCAGATATGGGATTTTAACTTG
GGACAGTCACGGGGACCTGAGGATACGAGTCCGAGTGAAGCTGCATATGTAGGGAAAGGT
GCTGCTTCTTCATTACAAATCAACAATTTTGTGACCATATGAATGAACTTGTTCCTACT
AATGTGAAAGGTGTCAAAGAGATTAAAAGGATGACTACAAGCGATCAACTTCAGGCCAG
GTACAACCAACAAAATCTGAGAGCAACAATCGTCCAATTACCTTTGGCTCTGAGAAAGGT
TCGAACTCCTCCAGTGACTTGCAATTTACAGAGCATATTGCTGGAACCTAGTTGTAAGACC
ACAAGACTAGTTGCAACTAAGGCTGATCTGGAGCGGCTGGCTCAGAACAGAGGAGATGCA
ATGCAGCGTTACAAGGAAAAGAGGAAGACACGAGATATGATAAGACCATAAGGTATGAA
TCGAGGAAGGCAAGAGCTGACACTAGGTTGCGTGTGAGGCGAGATTGTGAAAGCTAGT
GAAGCTCCTTACCCTTAACCTTAAGTTTTTTTACATAGGCTTCTTTTAGCTACAACTT
AGTTACTTTTTTTACTCCACTGCCTCATAAATGTACAGACCGGTCTCGTTTCATCTGGCC

GCCCTTCTTGTGTTTATTGCCTTATCTGGCCCTTTTATGTACCTTGAATCTTATCTAGTT
TAAAAAAGATTGTAACTTCTAGAAAACCATATTCTGTTGACAGTATATACATGCTCTATC
CAAGCAAAAA

>G1337 Amino Acid Sequence (domain in AA coordinates: 9-75)
MSSSERVPCDFCGERAVLFCRADTAKLCLPCDQVHTANLLSRKHVRSQICDNCNEPV
SVRCFTDNLILCQECDDVHSGSCSVSDAHVRSVEGFGSCPSALELAALWGLDLEQGRKD
EENQVPMAMMMDNFMQLDSWVLGSNELIVPSDITFKRGSGSSCGRYKQVLCKQLEE
LLKSGVVGGDGDGDRDRDCDREGACDGDGDGEAGEGLMVPFEMSERLKWSDVEEINGGG
GGGVNQWNATTNPSSGQSSQIWDNFNLGQSRGPEDTSRVEAAYVGKGAASSFTINNFD
HMNETCSTNVKGVKEIKKDDYKRSTSGQVQPTKSESNNRPITFGSEKGSNSSDLHFTEH
IAGTSCKTTRLVATKADLERLAQNRGDAMQRYKEKRKTRRYDKTIRYESRKARADTRLRV
RGRFVKASEAPY*

>G1759 (110..700)
CGAGAAAAGGAAAAAATAGAAAAGAGAAAACGCTTAGTATCTCCGGCGACTTGAAC
CCAAACCTGAGGATCAAATTAGGGCACAAAGCCCTCTCGGAGAGAAGCCATGGGAAGAAA
AAAAGTAGAAATCAAGCGAATTGAGAACAAAAGTAGCCGACAAAGTCACTTCTCAAACG
TCGCAACGGTCTCATCGAGAAAGCTCGTCAGCTTTCTGTTCTCTGTGACGCATCCGTCGC
TCTTCTCGTCTCTCCGCCCTCCGGCAAGCTCTACAGCTTCTCCTCCGGCGATAACCTGGT
CAAGATCCTTGATCGATATGGGAAACAGCATGCTGATGATCTTAAAGCCTTGGATCATCA
GTCAAAGCTCTGAAGTATGGTTCACACTATGAGCTACTTGAAGCTTGTGGATAGCAAGCT
TGTGGGATCAAATGCAAAAATGTGAGTATCGATGCTCTTGTTCAACTGGAGGAACACCT
TGAGACTGCCCTCTCCGTGACTAGAGCAAGAAGACCGAAGTCACTGTTGAAGCTTGTGA
GAATCTTAAAGAAAAGGAGAAAATGCTGAAAGAAGAGAACAGGTTTGGCTAGCCAGAT
GGAGAATAATCATCATGTGGAGCAGAAGCTGAGATGGAGATGTCACCTGCTGGACAAAT
CTCCGACAACTCTCCGGTACTCTCCACTACTTAATTAGCCACCTTAAATCGGCGGTG
AAATCAAATCCAAAACATATATAATTATGAAGAAAAAATAAGATATGTAATTATT
CCGCTGATAAGGGCGAGCGTTTGTATATCTTAATACTCTCTCTTTGGCCAAGAGACTTTG
TGTGTGATACTTAAGTAGACGGAAGTCAATACTATCTGTTTAAAGACAAAAGGTTG
ATGAAGTTGTACCTTATTCGTGTGAGAAAAA

>G1759 Amino Acid Sequence (conserved domain in AA coordinates: 2-57)
MGRKKLEIKRIENKSSRQVTFSKRRNGLIEKARQLSVLCDASVALLVVSASGKLYSFSSG
DNLVKILDRYQKHADDLKALDHQSKALNYGSHYELLELVDSKLVGSNVKNVSIDALVQL
EEHLETALSVTRAKKTEMLKLVENLKEKEKMLKEENQVLASQMNHNHVGAEAEEMESP
AGQISDNLPVTLPLLN*

>G1804 (169..1497)
TATCTCTCTCTTTCTCAAACCTTTTCAGTCAAATTTCTCCGGCGGCTTTTAAACTATGTG
AAGGAGGAGAACCTCCATAACAAGAAGCGGATTCTCTCAGTTTTCGGCGGCGGAGGAAC
ACAAAGCCACCGGTTTTAGACACACAGATTTTCAATTTTCAAGTTGTTAAATGGTAAGTAGA
GAAACGAAGTTGACGTCAGAGCGAGAAGTAGAGTCCGTCATGGCGCAAGCGAGACATAAT
GGAGGAGGTGGTGGTGAGAATCATCCGTTTACTTCTTTGGGAAGACAATCCTCTATCTAC
TCATTGACCTTGACGAGTTCCAACATGCTTTATGTGAGAACGGCAAGAAGTTGGGTCC
ATGAACATGGACGAGTTTCTTGTCTCTATTTGGAACGAGAGGAGAATAATAACAATCAA
CAACAAGCAGCAGCAGCTGCAGGTTCACTTCTGTTCCGGCTAATCAATGGTTTCAAC
AACAACAATAACAATGGAGGCGAGGTTGGTGTGTTAGTGGTGGTTCTAGAGGC
AACGAAGATGCTAACAATAAGAGAGGGATAGCGAACGAGTCTAGTCTTCTCGACAAGGC
TCTTTGACACTTCCAGCTCCGCTTGTAGGAAGACTGTTGATGAGGTTTGGTCTGAGATA
CATAGAGGTGGTGGTAGCGGTAATGGAGGAGACGCAATGGACGTAGTAGTAGTAAT
GGACAGAACAATGCTCAGAACGGCGGTGAGACTGCGGCTAGACAACCGACTTTTGGAGAG
ATGACACTTGAGGATTTCTTGTGTAAGGCTGGTGTGTTAGAGAACATCCCACTAATCCT
AAACCTAATCCAAACCCGAACCAAAACCAAAACCGTCTAGTGTAATACCCGAGCTGCA
CAGCAACAGCTTTATGGTGTGTTTCAAGGAACCGGTGATCCTTCATTCGCGGTCAAGCT
ATGGGTGTGGGTGACCCATCAGGTTATGCTAAAAGGACAGGAGGAGGAGGTATCAGCAG
GCGCCACCAGTTCAAGCAGGTGTTGCTATGGAGGTGGCGTTGGGTTTGGAGCGGTGGA
CAGCAATGGGAATGGTTGGACCGTTAAGCCCGGTGCTTCAGATGGATTAGGACATGGA
CAAGTGGATAACATAGGAGGTAGTATGGAGTAGATATGGGAGGGCTAAGGGGAAGGAAA
AGAGTAGTGGATGGTCCAGTGGAGAAAGTAGTGGAGAGAAGACAGAGGAGGATGATCAAG
AACCGCGAGTCTGCTGCTAGATCTAGAGCAAGAAAACAAGCATATACAGTGAATTGGAA

GCTGAACTTAACCAGTTGAAAGAAGAGAATGCGCAGCTAAAAACATGCATTGGCGGAGTTG
GAGAGGAAGAGGAAGCAACAGTATTTTGAGAGTTTGAAGTCAAGGGCACAAACCGAAATTG
CCGAAATCGAACGGGAGATTGCGGACATTGATGAGGAACCGAGTTGTCCACTCTAAACA
AACAATAGGAAGATGGAGAAGAAGTCGGAGACAGAACGAGGGAAAAACTGATGATTTTCT
ACGTTGTTGTTTTGTCTTTGAGGAATGAGGTTATAGAATCTTTATACTTTGATGTTTTCT
GTGTTGGTAGGAGGAACACCATCTGATCTGCTTTACTAGTGTTCCCTGTGAACAAAGAAA
GTGATTCTGTGTTTCAACATCATCAATCTTTGAAA

>G1804 Amino Acid Sequence (domain in AA coordinates: 357-407)

MVTRETKLTSEREVESMAQARHNGGGGGENHPFTSLGRQSSIYSLTLDEFQHALCENGK
NFGSMNMDEFLVSIWNAEENNNNQOAAAAAGSHSV PANHNGFNNNNNNGGEGVGVFSG
GSRGNEDANNKRGIANESSLPQGSLLPAPLCRKTVDEVWSEIHRGGSGNGGDSNGRS
SSNGQNNNAQNGGETAARQPTFGEMTLEDFLVKAGVVRHPTNPKPNPNQNPSSVI
PAAAQQQLYGVFQGTGDPSPFGQAMGVGDPSPGYAKRTGGGGYQQAPPVQAGVCYGGGVGF
GAGGQMGVMVGPLSPVSSDGLGHGQVDNIGGQYGVDMGGLRGRKRVVDGPVEKVVERRQR
RMKKNRESAARSARKQAYTVELEAELNQLKEENAQLKHALAELEKRRKQYFESLKSRA
QPKLPKSNGLRLTLMRNPSCPL*

>G207 (16..930)

aaaagatctgtttcaatggcggatcgtgttaagggtccatggagtcagaagaagatgag
cagctacgaaggatgggttgagaaatacggaccgaggaattggtctgcgattagcaaatcg
attccaggtcgactcgtgtaaatcgtgtagattacgttggtgtaatcagttatctccggag
gttgagcatcgtcctttctcgccggaggaagatgagactattgtaaccgcccgtgctcag
tttggttaacaagtggcgacgattgctcgtcttctaaccggtcgtacggataacgccgtt
aaaaatcactggaactctacgcttaagaggaaatgcagcggaggtgtggcggttacgacg
gtgacggagacggaggaagatcaggatcgccgaagaagaggagatctgttagctttgat
cctgcttttgctccggtggatactggattgtacatgagtcctgagagtcctaaccggaatc
gatgtagtgattctagcacgattccgtcaccgtcgtctcctggtgctcagctgtttaaa
ccaatgccgatttccggcggttttacgggtggttcgcgagccgttacgggttgaaatgtct
tcgtcttcggaggatccacctacttcgttgagtttgtcactacctggagctgagaacacg
agttcgagccataacaataacaacaacgcgttgatgtttccgagatttgagagtcagatg
aagattaatgtagaggagagaggaggaggaggagaaggacgtagaggtgagtttatgacg
gtggtgcaggagatgataaaagctgaagtgaggagttacatggcggaaatgcagaaaaca
agtgggtggattcgtcgtcggaggtttatagcaatccggcggaatggtggttttagggat
tgtggagtaataaacacctaaaggttgagtagtttgggttaggggttaaaacttgaaatcgat
tggggattttcaagagcattcatttttgggggttatggtaaaattaaaaacaaaaacaaa
atgtacagaggaattaaaatttctatggaataatcttaaatctcaaatatttgttacttg
ttttggtgattcataacccaaatcaaa

>G207 Amino Acid Sequence (domain in AA coordinates: 6-106)

MADRVKGPWSQEEDQLRRMVEKYGPRNWSAISKSIIPGRSGKSCRLRWCNQLSPEVEHRP
FSPPEEDETIVTARAQFGNKWATIARLLNGRTDNVKNHWNSTLKRKCSGGVAVTTVTETE
EDQDRPKRRRSVSFDPAPFVDTGLYMSPEPSPNGIDVSDSSTIPSPSSPVAQLFKPMPIS
GGFTVVPQPLPVEMSSSEDPPTSLSLSLPGAENTSSSHNNNNNALMFPFRFESQMKINVE
ERGGGEGRRGEFMTVVQEMIKAEVRSYMAEMQKTSGGFVVGGLYESGGNGGFRDCGVIT
PKVE*

>G218 (1..1182)

ATGGAGGCAGAGATCGTGAGACGATCGGAGGTAACGGGATTAAGAAGGGAGGTGGAAGAA
TCGTCAATTGGTAGAGGAGATTGCGATGCGTATGCGGCGATGTGGGAGAAGATGCGGCA
GGGTTCTGTGGGACGAGCGGGAGAGGAAGAAGAGATCGAGTTAAAGGGCCGTGGTCAAG
GAGGAGGATGATGTGTTGAGTGAGCTCGTTAAGAGGTTGGGAGCGAGGAATTGGAGTTTT
ATCGCTCGGAGTATTCCTGGTCTGTTCAAGCAAGTCTTGTCTCTTCTGTTGGTGTAATCAG
CTCAATCCAAATCTTATACGCAATTCATTACTGAGGTAGAGGATCAGGCTATCATCGCA
GCACATGCCATCCACGGAAACAAATGGGCTGTTATCGCGAAGCTCCTCCCCGGAAGACA
GATAATGCTATCAAGAACCCTGGAACCTCTGCTTTAAGACGTCGATTATAGACTTTGAA
AAGCCCAAGAAATATAGGAACCTGGAAGCTTGGTCTGTGGATGATTCTGGATTGACAGAACG
ACAACAGTAGCCTCATCAGAAGAACTTTATCTTCAGGCGGTGGTTGCCATGTAACACTACT
CCAATTGTATCTCCAGAAGGCAAAGAAGCTACCACCTCCATGGAAATGTCTGAAGAACAA
TGCGTAGAGAAAACAAACGAGAAGGTATTTCTAGGCAAGATGATAAGGATCCTCCAACG
CTTTTCGCCCCAGTGCTCGGCTCAGTTCTTTTAAATGCTTGCAATCACATGGAAGGATCA

CCCTCTCCACATATACAAGACCAAAATCAGCTCCAATCATCTAAACAAGACGCAGCAATG
CTAAGATTGCTTGAAGGAGCTTACAGCGAACGGTTTGTGCCTCAAACATGTGGAGGTGGT
TGTTCAGCAACAATCCCGATGGCAGTTTTTCAGCAAGAATCATTGTTGGGTCCAGAGTTT
GTGGATTACTTAGACTCACCAACGTTTCCGAGTTCGAACTAGCTGCTATAGCAACGGAA
ATAGGCAGCCTCGCTTGGCTGAGAAGCGGTTTAGAGAGTAGCAGCGTGAGGGTGATGGAA
GACGCAGTTGGTCGGTTAAGGCCTCAAGGCTCCAGGGGTTCATCGAGATCATTATCTTGTA
TCTGAACAGGGGACGAACATAACCAATGTCCTGTCCACATAA

>G218 Amino Acid Sequence (domain in AA coordinates: TBD)
MEAEIVRRSEVTGLRREVEESSIGRGDCDGDGDVGEDAAGFVGTSGRGRDRVKGPWSK
EEDDLVSELVKRLGARNWSFIARSIPGRSGKSCRLRWCNQLNPNLIRNSFTEVEDQAIIA
AHAIHGNKWAVIAKLLPGRTDNAIKNHWNSALRRRFIDFEKAKNIGTSLVVDSDSGFDRT
TTVASSEETLSGGGGCHVTTPIVSPEGKEATTSMEMSEEQVEKTNGEGISRQDDKDPPT
LFRPVPRLSSFNACNHMEGSPSPHIQDQNLQSSKQDAAMLRLLEGAYSERFVPQTCCGG
CCSNPDGSGFQGESLLGPEFVDYLDSPTFPSSELAIAIEIGSLAWLRSGLESSSVRVME
DAVGRRLRPQGSRGHRDHYLVSEQGTNITNVLST*

>G241 (46..867)
GAAAAAATTTCAACTTCTTTTATCAGCAATCACAAATCAAAGAGATGGGAAGAGCTCCA
TGCTGTGAGAAGATGGGGTTGAAGAGAGGACCATGGACACCTGAAGAAGATCAAATCTTG
GTCTCTTTTATCCTCAACCATGGACATAGTAAGTGGCGAGCCCTCCCTAAGCAAGCTGGT
CTTTTGAGATGTGGAAGAGCTGTAGACTTAGGTGGATGAACATTTAAAGCCTGATATT
AAACGTGGCAATTTACCAAAGAAGAGGAAGATGCTATCATCAGCTTACACCAATACTT
GGCAATAGATGGTCAGCGATTGCAGCAAACTGCCGTGAAGAACCGATAACGAGATCAAG
AACGTATGGCACACTCACTTGAAGAAGAGACTCGAAGATTATCAACCAGCTAAACCTAAG
ACCAGCAACAAAAGAAGGGTACTAAACCAAAATCTGAATCCGTAATAACGAGCTCGAAC
AGTACTAGAGCGAATCGGAGCTAGCAGATTATCAAAACCTTCTGGAGAAAGCTTATTT
TCGACATCGCCTTCGACAAAGTGAGGTTCTTCGATGACACTCATAAGCCACGACGGCTAT
AGCAACGAGATTAATATGGATAACAAACCGGGAGATATCAGTACTATCGATCAAGAATGT
GTTTCTTTTCGAACTTTTGGTGCGGATATCGATGAAAGCTTCTGGAAAGAGACTGTAT
AGCCAAGATGAACACAACTACGTATCGAATGACCTAGAAGTCGCTGGTTTAGTTGAGATA
CAACAAGAGTTTCAAACTTGGGCTCCGCTAATAATGAGATGATTTTTCAGAGTGAGATG
GAACCTCTGGTTCGATGTATTGGCTAGAACCGGCGGGGAACAAGATCTCTTAGCCGGGCT
CTAGTTAACATGTTTGGAGAGTAAAGTGAAATGGTGCAAATTAGTTAAGGCTAAGAAATT
CAAAAGCTTTGTTTACCGAGAAAAAACAACACTCTAACTCTTGATGTGATGTAGTTAGT
GTATTAATTAGAGGCTGCGTTTTCAA

>G241 Amino Acid Sequence (domain in AA coordinates: 14-114)
MGRAPCCEKMGLKRPWTPPEEDQILVSVFILNHGHSNWRALPKQAGLLRCGKSCRLRWMNY
LKPDIKGNFTKEEDAIISLHQILGNRWSAIAKLPGRDNEIKNVWHTLKKRLEDYQ
PAKPKTSNKKKGTKPKSESVITSSNSTRSESELADSSNPSSGESLFTSPSTSEVSSMTLI
SHDGYSEINMDNKPDIIDTIDQECVSFETFGADIDESFWKETLYSQDEHNYVSNDDLEVA
GLVEIQQEFQNLGSANNEMIFDSEMELLVRCIG*

>G254 (15..923)
CGATTTTCGAGCTCTATGGTGTCCGTAAACCCCTAGACCTAAGGGTTTTCCAGTTTTTCGATT
CCTCGAATATGAGTTTACCAAGCTCCGATGGATTGTTTCGATTCCGGCCACGGGACGGA
CCAGTACGGTGTCTGTTTTCTGAGGATCCGACGACGAAGATTCCGAAGCCGTACACAATCA
AGAAGTCGAGAGAGAATTGGACAGATCAAGAGCAGGATAAATTTCTAGAAGCTCTTCACT
TATTCGATAGGGATTGGAAGAAAATAGAAGCCTTTGTTGGATCAAAAACAGTAGTTCAGA
TACGAAGCCACGCTCAGAAATACTTTCTCAAAGTTCAGAAGAGTGGTGCTAACGAACATC
TTCCACTTCTCGACCTAAGAGGAAAGCGAGTCATCCTTATCCTATAAAGGCTCCTAAAA
ATGTTGCTTATACCTCTCTCCGCTCTTCGAGTACATTACCGTTGCTTGAGCCTGGTTATT
TGTATAGCTCTGATTGGAAGTCATTGATGGGAAACAGGCTGTTTGTGCATCTACCTCTT
CTTCGTGGAATCATGAATCGACAAATCTGCCAAAACCGGTGATTGAAGAGGAACCGGGAG
TCTCGGCCACGCTCCTCTCCCAAATAATCGCTGCAGACAGGAAGATACAGAGAGGGTAC
GAGCAGTGACAAAGCCAAATAACGAAGAAAGTTGTGAAAGCCACATAGAGTGATGCCGA
ATTTTGCTGAAGTTTACAGCTTCATTGGAAGTGTCTTCGATCCCAACACATCAGGCCACC
TCCAGAGATTAAAGCAGATGGATCCAATAAATATGGAACGGTCTTTTACTGATGCAAAA
ACCTGTCTGTAAATCTGACAAAGTCCCGAGTTTGCAGAGCAAGGAGGTTGATATCATCAT
ACAGCGCTAAAGCTTTGAAATAGAGATAGAATAAAACAATAATGTACCTTATGTGATC

AAGAGACAATCATCCAAGGTCTGTATGCATTGCTTGGATTTAGGCCTCGTGTCTCACTA
CAGGAGCAGAACCAATCGCAAAGACTCTTAGATGGCTACTGAGTTGTGGTTTTATGTCT
CTGTAAGTCGCGGTGGAGCACACGTGTTTGTCTGTCTTGTGTATGTGTGTATAGATAAT
ACAAGGTTTTGTCAGAGTAAGGTCACAGTTAGCTGCAAGTGAGTTGGATCAATCTTAAGA
TTAAACCCTTGAGAGTGAGTGTCCAAAGAGACTGTGTAATATTGTTTGGCGGTCAGCAG
AAGAGTTTTGAAGTGCACATCCAGTTAGTGATAACACGGTTGAAGAAAAGGTAAGGTTAC
AAGTTTAGTTTTGAATAATTGTATACTCAAAAAATATGAATGTATAAAGAATAATCACTT
GAGTCGCCTTA

>G254 Amino Acid Sequence (domain in AA coordinates: 62-106)
MVSVNPRPKGFPVFDSSNMSLPSSDGFSGIPATGRSTVSFSEDPTTKIRKPYTIKKSRE
NWTQDQHDKFLEALHFLDRDWKKIEAFVGSKTIVQIRSHAQKYFLKVQKSGANEHLPLPR
PKRKASHPYPIKAPKNVAYTSLPSSSTLPLEPGYLYSSDSKSLMGNQAVCASTSSSWNH
ESTNLPKPVIEEPPGVSATAPLPNNRCRQEDTERVRAVTKPNNEESCEKPHRVMPNFAEV
YSFIGSVFDPNTSGHLQRLKQMDPINMETVLLLMQNL SVNLTSPFAEQRLISSYSAKA
LK*

>G26 (73..729)

TTGGCTTGTAACCAACCCATCTTTGACTTCAAAAATAAAAATAAATCATAATTGA
CATCATCGGATAATGCATAGCGGGAAGAGACCTCTATCACCAGAATCAATGGCCGGAAAT
AGAGAAGAGAAAAAGAGTTGTGTGTTGCTCAACTTTGTCGGAATCTGATGTGTCTGAT
TTTGTCTCTGAAGTCACTGCTCAACCCATCCCATCATCCATTGATGATCAATCTTCGTCG
CTTACTCTTCAAGAAAAAGTAAGTCAAGGCAACGAACTACAGAGGCGTGAGGCAAGA
CCGTGGGGAAAAATGGGCGGCTGAGATTCTGTGACCCGAACAAGGCAGCTCGTGTGTGGCTT
GGGACGTTTCGACACTGCAGAAGAAGCCGCTTAGCGTATGATAAAGCTGCATTTGAGTTT
AGAGGTCACAAGGCCAAGCTTAACCTTCCCGAGCATATTCGTGTCAACCCCTACTCAACTC
TATCCATCGCCCGCTACTTCCCATGATCGCATTATCGTGACACCACCTAGTCCACCTCCA
CCAATTGCTCCTGACATACTTCTTGATCAATATGGCCACTTCAATCTCGAAGTAGTGAT
TCCAGTGCCCAACTTGTCCATGAATATGCTGTCTTCTTCTGCTTTCATCTTGAATCATCAA
GGGCTAAGACCAAAATTTGGAGGATGGTGAAGACGTAAGAACATTAGTATCCACAAACGA
CGAAAATAACATGTTAATGGCATAAATATCTTCTGTCGAAGTTATCAAACGCATTGACC
TCCGGCTTTGATCATTTTAGGCGCTTAATCTCTTTACGACTTCATTTTGGTAGTCTTTAA
AGAGTCTATGGAGTGGATTTAGCTAGGAATCAGGCCTTATGGATGAAAAATATATAAATT
TTGAACATGACTATGCAAGATGGGATGAAGACTACTTAGCTTGAAAAACGTCCTGATAG
GTCATGACGACTATATCCACAGAAGATGACCGACGGAGACAACAACATGCCTCACCTGAT
CGACCGATCAAAATGAGATAATGTGTTGACCGGACCGGTCGGATCAGGTTGGGTGAGTAT
ATCA

>G26 Amino Acid Sequence (domain in AA coordinates: 67-134)
MHSGKRPLSPESMAGNREEKKELCCSTLSESDVSDVSELTGQIPSSIDDQSSSLTLQ
EKSNSRQRNYRQVRQRPWGKWAABEIRDPNKAARVWLGTDFDTAEAAALAYDKAAFEFRGHK
AKLNFPEHIRVNPTQLYPSPATSHDRIIVTPPSPPPPIAPDILLDQYGHFQSRSSDSSAN
LSMNLSSSSSLNHQGLRPNLEDGENVKNISIHKRRK*

>G263 (48..902)

TTTTTAGTTTTATTTTCTGTGGTAAAATAAAAAAGTTTCGCCGGAGATGACGGCTGTGA
CGGCGGCGCAAAGATCAGTTCCGGCGCCGTTTTTAAGCAAAACGTATCAGCTAGTTGATG
ATCATAGCACAGACGACGTCGTTTCATGGAACGAAGAAGGAACAGCTTTTGTCTGTGGA
AAACAGCAGAGTTTGTAAAGATCTTCTTCTCAATACTTCAAGCATAATAATTCTCAA
GCTTCATTCTGTCAGCTCAACACTTACGGATTTTCGTAAAACTGTACCGGATAAATGGGAAT
TTGCAAAACGATTATTTCCGGAGAGGCGGGGAGGATCTGTTGACGGACATACGACGGCGTA
AATCGGTGATTGCTTCAACGGCGGGGAAATGTGTTGTTGTTGTTGCTTCTGAGTCTA
ATTCTGTTGGTGGTGTGATGATCACGGTTCAAGCTCCACGTCATACCCGGTTCGTGGAAGA
ATCCTGGTTCCGTTGGAGAATGTTGCTGATTTATCAGGAGAGAACGAGAAGCTTAAAC
GTGAAAACAATAACTTGAGCTCGGAGCTCGCGGCGGGAAGAAGCAGCGGATGAGCTAG
TGACGTTCTTGACGGGTCTCTGAAAGTAAGACCGGAACAAATCGATAAAATGATCAAAG
GAGGGAAATTTAAACCGGTGGAGTCTGACGAAGAGAGTGAGTGCGAAGGTTGCGACGGCG
GCGGAGGAGCAGAGGAGGGGGTAGGTGAAGGATTGAAATGTTTGGGGTGTGGTTGAAAG
GAGAGAGAAAAAAGAGGACCGGGATGAAAAGAATTATGTGGTGTGAGTGGGTCCCGTATGA
CGGAAATAAAGAACGTGGACTTTACGCGCGGTTGTGGAAAAGCAGCAAAGTCTGCAACT
AAAAAAGAGTAGAAGACTGTTCAAACAGCGGTGTGACACGTCATCAGCAGCAGCAAAA

AAATGATTTAAAAAACTATTTTTTCCGTAAGGAAGAAAAGTTATTTTATGTTTAAAA
AGGTGAAGAAGGTCCAGAAGGATCAACGCAAATATATAAATGGATTTTCATGTATTATAT
AATTTAATTAGTGTATTAAGAAAA

>G263 Amino Acid Sequence (domain in AA coordinates: TBD)
MTAVTAAQSRVPAFLSKTYQLVDDHSTDDVVSWEETAFVVKTAFAKDLLPQYFKH
NNFSSFIQLNTYGFRTKTPDKWEFANDYFRRGGEDLLDIRRRKSVIASTAGKCVVVG
PSESNSGGDDHGSSTSSPGSSKNPGSVENMVADLSENEKLKRENNNLSSSELAAAKKQ
RDELVTFLTGHKLKVRPEQIDKMIKGGKFKPVESDEESECEGCDGGGGAEEGVGEGLKLF
VWLKGERKKRDRDEKNYVVSGSRMTEIKNVDFHAPLWKSSKVCN*

>G308 (196..1794)
AGTAATTTAGTTTTTTTTTTTTTTTACAATTTATTTTGTATTAGAAAGTGGTAGTGG
AGTGAaaaaaCAAATCCTAAGCAGTCCTAACCGATCCCCGAAGCTAAAGATTCTTCACCT
TCCCAAATAAAGCAAAACCTAGATCCGACATTGAAGGAAAAACCTTTTAGATCCATCTCT
GAAAAAAACCCAACCATGAAGAGAGATCATCATCATCATCAAGATAAGAAGACTATG
ATGATGAATGAAGAAGACGACGGTAACGGCATGGATGAGCTTCTAGCTGTTCTTGGTTAC
AAGGTTAGGTCATCGGAAATGGCTGATGTTGCTCAGAACTCGAGCAGCTTGAAGTTATG
ATGTCTAATGTTCAAGAAGACGATCTTCTCAACTCGCTACTGAGACTGTTCACTATAAT
CCGGCGGAGCTTTACACGTGGCTTGATTCTATGCTCACCAGCCTTAATCCTCCGTCGTCT
AACGCCGAGTACGATCTTAAAGCTATCCCGGTGACGCGATTCTCAATCAGTTCGCTATC
GATTCCGCTTCTTCGCTTAACCAAGGCGGCGGAGGAGATACGTATACTACAAACAAGCGG
TTGAAATGCTCAAACGGCGTCGTGGAAACCACCACAGCGACGGCTGAGTCAACTCGGCAT
GTTGCTCCTGGTTGACTCGCAGGAGAACGGTGTGCGTCTCGTTACCGCGCTTTTGGCTTGC
GCTGAAGCTGTTTCAGAAGGAGAATCTGACTGTGGCGGAAGCTCTGGTGAAGCAAATCGGA
TTCTTAGCTGTTTCTCAAATCGGAGCTATGAGACAAGTCGCTACTTACTTCGCGGAAGCT
CTCGCGCGGCGGATTTACCGTCTCTCTCCGTCGCAGAGTCCAATCGACCACTCTCTCTCC
GATACTCTTCAGATGCACTTCTACGAGACTTGTCTTATCTCAAGTTGCTCACTTCACG
GCGAATCAAGCGATTCTCGAAGCTTTCAAGGGAAGAAAAGAGTTTCATGTCAATGATTTT
TCTATGAGTCAAGGTCTTCAATGGCCGCGCTTATGCAGGCTCTTGCCTTCGACCTGGT
GGTCTCTCTGTTTTCCGGTTAACCAGAAATGGTCCACCGGCACCGGATAATTTGATTAT
CTTCATGAAGTTGGGTGTAAGCTGGCTCAATTTAGCTGAGGCGATTACGTTGAGTTTGA
TACAGAGGATTTGTGGCTAACACTTTAGCTGATCTTGATGCTTCGATGCTTGAGCTTGA
CCAAGTGAGATTGAATCTGTTGCGGTTAACTCTGTTTTTCGAGCTTCACAAGCTCTTGGGA
CGACCTGGTGCGATCGATAAGGTTCTTGGTGTGGTGAATCAGATTAAACCGGAGATTTTC
ACTGTGGTTGAGCAGGAATCGAACATAATAGTCCGATTTTCTTAGATCGGTTTACTGAG
TCGTTGCATTATTACTCGACGTTGTTGACTCGTTGGAAGGTGTACCGAGTGGTCAAGAC
AAGGTCATGTGCGGAGGTTTACTTGGGTAAACAGATCTGCAACGTTGTGGCTTGTGATGGA
CCTGACCGAGTTGAGCGTCATGAAACGTTGAGTCAGTGAGGAACCGGTTGCGGCTCTGCT
GGGTTTGGCGCTGCACATATTGGTTTGAATGCGTTTAAAGCAAGCGAGTATGCTTTTGGCT
CTGTTCAACGGCGGTGAGGGTTATCGGGTGGAGGAGAGTGACGGCTGTCTCATGTTGGGT
TGGCACACACGACCGCTCATAGCCACCTCGGCTTGAAACTCTCCACCAATTAGATGGTG
GCTCAATGAATTGATCTGTTGAACCGGTTATGATGATAGATTCCGACCGAAGCCAAACT
AAATCCTACTGTTTTTCCCTTTGTCACTTGTTAAGATCTTATCTTTCATTATATTAGGTA
ATTGAAAAATTTTAATCTCGCCTAAATTACT

>G308 Amino Acid Sequence (domain in AA coordinates: 270-274)
MKRDHHHHQDKKMTMMNEEDDGNMDELLAVLGYKVRSEEMADVAQKLEQLEVMSNVQ
EDDLSQLATETVHYNPAELYTWLDSMLTDLNPPSSNAEYDLKAIPGDAILNQFAIDSASS
SNQGGGGDYTTNKRLLKCSNGVVETTTTATAESTRHVVLDVDSQENGVRVHALLACAEAVQ
KENLTVAEALVKQIGFLAVSQIGAMRQVATYFAEALARRIYRLSPSQSPIDHSLSDLQOM
HFYETCPYLKFAHFTANQAILEAFQGGKRVHVIDFMSQGLQWPALMQALALRPGGPPVF
RLTGIGPPAPDNFDYLFHEVGCKLAHLAEAIHVEFEYRGFVANTLADLDASMLELRPSEIE
SVAVNSVFEHLKHLGRPGAIDKVLGVVNQIKPEIFTVVEQESNHNSPIFLDRFTESLHY
STLFDSLEGVPSGQDKVMSEVYLKGQICNVVACDGPDRVERHETLSQWRNRFSGAGFAAA
HIGSNAFKQASMLLALFNGGEGYRVEESDGCLMLGWHTRLIATSANKLSTN*
>G38 (149..1156)
GAGGAAACTCGAAAAAGCTACACACAAGAAGAAGAAAAGATACGAGCAAGAAGACT
AAACACGAAAGCGATTTATCAACTCGAAGGAAGAGACTTTGATTTTCAAATTCGTCCCC
TATAGATTGTGTTGTTTCTGGGAAGGAGATGGCAGTTTATGATCAGAGTGGAGATAGAAA

CAGAACACAAATTGATACATCGAGGAAAAGGAAATCTAGAAGTAGAGGTGACGGTACTAC
TGTGGCTGAGAGATTAAAGAGATGGAAAGAGTATAACGAGACCGTAGAAGAAGTTTCTAC
CAAGAAGAGGAAAGTACCTGCGAAAGGGTCGAAGAAGGGTTGTATGAAAGGTAAAGGAGG
ACCAGAGAATAGCCGATGTAGTTTCAGAGGAGTTAGGCAAAGGATTTGGGGTAAATGGGT
TGCTGAGATCAGAGAGCCTAATCGAGGTAGCAGGCTTTGGCTTGGTACTTTCCCTACTGC
TCAAGAAGCTGCTTCTGCTTATGATGAGGCTGCTAAAGCTATGTATGGTCCTTTGGCTCG
TCTTAATTTCCCTCGGTCTGATGCGTCTGAGGTTACGAGTACCTCAAGTCAGTCTGAGGT
GTGTACTGTTGAGACTCCTGGTTGTGTTTCATGTGAAAACAGAGGATCCAGATTGTGAATC
TAAACCCCTTCTCCGGTGGAGTGGAGCCGATGTATTGTCTGGAGAATGGTGCAGGAGAT
GAAGAGAGGTGTTAAAGCGGATAAGCATTGGCTGAGCGAGTTTGAACATAACTATTGGAG
TGATATTCTGAAAGAGAAAGAGAAACAGAAGGAGCAAGGGATTGTAGAAACCTGTGAGCA
ACAAACAGCAGGATTTCGCTATCTGTTGACAGACTATGGTTGGCCCAATGATGTGGATCAGAG
TCACTTGGATTCTTCAGACATGTTTGATGTGATGAGCTTCTACGTGACCTAAATGGCGA
CGATGTGTTTGCAGGCTTAAATCAGGACCGGTACCCGGGGAACAGTGTTCGCAACGGTTC
ATACAGGCCCCGAGAGTCAACAAAGTGGTTTTGATCCGCTACAAAGCCTCAACTACGGAAT
ACCTCCGTTTCAGTTCGAGGGAAAGGATGGTAATGGATTCTTCGACGACTTGAGTTACTT
GGATCTGGAGAATAAACAACAATATGAAGCTTTTGGATTGTATTTGCCCTAATC
CCACAACGACTGTTGATTCTCTATCCGAGTTTTAGTGATATAGAGAACTACAGAACACGT
TTTTCTTGTATATAAAGGTGAAGTGTATATATCGAAACAGTGATATGACAATAGAGAAGA
CAACTATAGTTTGTAGTCTGCTTCTCTTAAGTTGTTCTTTAGATATGTTTTATGTTTTG
TAACAACAGGAATGAATAATACACACTTGTGAAGCTTTAAAAAAAAAAAAAAAAAAAAA
>G38 Amino Acid Sequence (domain in AA coordinates: 76-143)
MAVYDQSGDRNRQTIDTSRKRKSRSGDGTVAERLKRWKEYNETVEEVSTKKRKVPAKG
SKKGCMKKGKGPENSRCSFRGVRQRIWGKWVAEIREPNRGSRLWLGTFFPTAQEAASAYDE
AAKAMYGPLARLNFPRSDASEVTSSTSSQSEVCTVETPGCVHVKTEDPDCESKPFSGGVEP
MYCLENGAEEMKRGVKADKHWLSEFEHNYWSDILKEKEKQKEQIVETCQQQQQDLSVA
DYGWPNDVDQSHLSDSDFDVELLRDLNGDDVFAGLNQDRYPGNSVANGSYRPESQQSG
FDPLQSLNYGIPPFQLEGKDGNGFDDLSYLDLEN*
>G43 (38..643)
CTCCTGTCTGTCTAAAGAAAAAGAGAGAGGAAGAAATGGAGACTTTTGAGGAAAGCTC
TGATTTGGATGTTATACAGAAACATCTATTTGAAGACTTGATGATCCCTGATGGTTTCAT
TGAAGATTTTGCTTTTGATGATCTGCTTTTGTCTCCGACTCTGGTCTCTAGAACCCTT
TAACCCAGTTTCCGAAACTGGAAACCTAGTTTCACCTGTTCTTGATCCAGATTCTATGTCCA
AGAGATTCTGCAAATGGAAGCAGAATCATCATCATCATCAACAACAACGTCACCTGA
GGTTGAGACTGTCTCAAAACCGGAAAAAAACAAAGAGGTTTGAAGAAACGAGACATTACAG
AGGCGTGAGAAGGAGGCCATGGGGGAAATTTGCAGCAGAGATTGAGATCCGGCAAAGAA
AGGATCCAGATTGTTGTTAGGCACTTTTGAGAGTGATATTGATGTGCAAGGGCTTACGA
CTATGCAGCTTTTAAGCTCAGGGGAAGAAAAGCTGTTCTCAACTTTCCTTTGGATGCCGG
AAAGTATGATGCTCCGGTCAATTCATGCCGAAAAAGGAGGAGAACCGATGTACCACAGCC
TCAAGGAACAACAACAAGTACTTCATCATCGTCATCAAACTAATGGGGGAATAGTGATGT
TTAATTAGTATATATAGGTTAATATCTTAAGTATGTGAAGCATCATGTATAGAGCCAAGA
ACCTGTTAGACTAGTGTACTGAAAAGAACTCTTGCAAAATATGTACTAAAGAGTTCCTGT
AACAATGGAACCTCTGCGTTTCTCTTGTCTTAAAGAGCTTAAGGTCTAGAAACAAAGT
TCTTGTCTTTCGGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAA
>G43 Amino Acid Sequence (domain in AA coordinates: 104-172)
METFEESDLVVIQKHLFEDLMIPDGFIEDFVFDDTAFVSGLSLEFPNPNVPKLEPSSPV
LDPDSYVQEIQLMEAESSSSSTSTTSPEVETVSNRKKTKRFEETRHYRGVRRRPWGKFAA
EIRDPKKGSRWLGTFSDDIDAARAYDYAAFKLGRKAVLNFLDAGKYDAPVNSCRKR
RRTDVPQPQGTSTSTSSSSN*
>G536 (1..768)
ATGTCGACAAGGGAAGAGAATGTTTACATGGCGAAATTAGCCGAACAAGCTGAACGTTAC
GAAGAAATGGTTGAATTCATGGAGAAAGTTGCGAAACTGTTGATGTTGAGGAACCTTCA
GTTGAAGAGAGGAATCTTCTCTGTTGCTTACAAGAACGTGATTGGAGCGAGAAGAGCT
TCGTGGAGAATCATTTCTTCGATTGAGCAGAAAGAAGAGAGCAAAGGGAACGAAGATCAT
GTTGCTATTATCAAGGATTACAGAGGAGAGATTGAATCCGAGCTTAGCAAAATCTGTGAT
GGGATTTTGAATGTTCTTGAAGCTCATCTTATTCCTTCTGCTTACCAGCTGAATCTAAA

GTGTTTTATCTTAAGATGAAGGGTGATTATCATAGGTATCTTGCTGAGTTTAAGGCTGGT
GCTGAAAGGAAAGAAGCTGCTGAAAGCACTTTGGTTGCTTACAAGTCTGCTTCCGACATT
GCCACTGCTGAGTTAGCTCCTACTACCCGATAAGGCTTGGTCTTGCACTCAACTTCTCT
GTGTTTTACTATGAAATCCTCAACTCGCCTGATCGTGCTTGCAAGCCTCGCAAAGCAGGCG
TTTGATGATGCAATCGCTGAGTTAGATACATTGGGTGAGGAATCATACAAGGACAGTACA
CTGATTATGCAGCTTCTTAGAGACAATCTCACTCTCTGGACTTCAGATATGACTGACGAA
GCAGGAGATGAGATTAAGGAGGCATCAAAGCCCGATGGTGCCGAGTAA
>G536 Amino Acid Sequence (domain in AA coordinates:226-233)
MSTREENVYMAKLAEQARYEEMVEFMKVAKTVDVEELSVBERNLLSVAYKNVIGARRA
SWRIISSIEQKEESKGNEDHVAIIKDYRGEIESELSKICDILNVLEAHLIPASPAESK
VFYLLKMGDGYHRYLAEFKAGAEKAAESTLVAYKSASDIATAELAPHTPIRLGLALNFS
VFYYEILNSPDRACSLAKQAFDDAIAELDTLGEESYKDSLIMQLLRDNLTLWTSMDTDE
AGDEIKEASKPDGAE*
>G567 (38..1273)
AAAAAGAAGAATCAGAAAGTAAAAAGAGAGCGAGCGATGAACAGTATCTTCTCCATTGA
CGATTTCTCCGATCCTTTCTGGGAACTCCTCCGATTCTCTCAATCCCGACTCTTCTAA
GCCTGTTACGGCGGATGAAGTTAGCCAGAGTCAACCGGAATGGACTTTCGAGATGTTTCT
CGAAGAGATTTCTCGTCGGCGGTGAGCTCTGAGCCACTTGGTAACAACAACAACGCGAT
CGTCGGTGTTTCTTCGGCGCAATCTCTTCTTCTGTTTCCGGACAGAATGATTTTCGAGGA
TGATAGTCGATTTCGTGATCGCGATTCCGGAAATTTGGATTGTGCTGCTCCCATGACGAC
GAAGACGGTGAATGTTGATTCGATGATTATCGTCGTGTTCTTAAGAACAAAGCTTGAGGC
TGAGTGCGCGACTGGTGTTTCTCTTCGGGTGGGTCTGTGAAGCCTGAAGATTGACTAG
TTCTCCAGAACTCAACTTCAACCAGTTCAATCCAGTCTCTTACTCAAGGAGAACTTGG
TGTTACTTCTTCTTACCAGCTGAGGTGAAAAAACTGGTGATCAATGAAGCAGGTAC
TAGTGATCCTCGAGAGAAATATTCTGATGACGAGGACCTTGATGAAGAGAATGAAACCAC
CGGTTCTTGAAGCCAGAGGACGTTAAAAAATCTAGAAGGATGCTGTCAAATCGTGAGTC
AGCTAGGCGATCTAGAAGGAGAAAGCAGGAGCAAACAAGTGACCTCGAAACACAGGTTAA
TGATCTAAAGGTGAGCATTCATCACTTCTTAAACAAGTGAACATGAATCACAAGTA
TGACGAGGCTGCTGTTGGCAATAGAACTAAAGGCTGACATTGAGACATTAAGAGCTAA
GGTGAAAAATGGCGGAAGAAACCGTGAAGAGAGTAACAGGAATGAATCCGATGCTTCTCGG
AAGATCAAGTGGACATAACAACAACAAGAGATGCCAATAACTGGTAACAACAGGATGGA
TTCTTCTAGCATTATTCCAGCTTATCAACCACACTCAAACCTAAACCATATGTCAAACCA
AAACATCGGGATCCCAACCATTCTACCTCCAAGACTCGGAAACAATTTCTGCTGCTCCTCC
ATCCCAAACAGCTCTCCCTTGACAGAGAATTAGAAATGGGCAAAATCACCATGTTACTCC
AAGCGCCAAACCGTATGGCTGGAATACCGAACCTCAGAACGATTGAGATGGCCGAAAAA
ATGCGTGGACTGATCAAACAAGAACGGGTTTCGCACTATATTAATGTCTATGCATCTGT
AATTTGTAAGTGTATTAAAGTTACGAATCATGAGAAAACATCTTGTGAAAATACAGTCTC
ATGGCTTATATATATATAAGCTCTGTCTTATAACATTACAAGATTCTTATTTGAGAAT
CGTCTTTCTATTTATAGCTAATAAAAAA
>G567 Amino Acid Sequence (domain in AA coordinates 210-270)
MNSIFSIDDFSDPFWETPIPLNPDSSKPVTADEVVSQSQPEWTFEMFLEISSAVSSEP
LGNNNNAIVGVSSAQSLPSVSGQNDFFEDDSRFRDRDSGNLDCAAPMTTKTVNVDSDDYRR
VLKNKLEAECATGVSRLRVGSVKPEDSTSSPETQLQPVQSSPLTQGELGVTSSLPAEVKKT
GVSMKQVTSGSSREYSDDEDLDEENETGSLKPEDVKKSRRMLSNRESARRSRRRKQEQT
SDLETQVNDLKGEHSSLLKQLSNMNHKYDEAAVGNRILKADIE TLRAKVKMAEETVKRVT
GMNPMLLGRSSGHNNNNRMPITGNRMDSIIIPAYQPHSNLNMNSNQNIPTILPPRL
GNNFAAPPSQTSSPLQIRNGQNHVTPSANPYGWNTPEQNDSAWPKKCVD*
>G580 (338..2275)
CAGTTATCTTCTTCTTCTTCTTCTGTTTTTAAATTTATTTTATAGAGAATTTTTTTTG
TTTTGCTTCCGATTGATTATTCCGGGAACGATGACTTCCGGGGAGTTCCCGGTGAG
ATGATAAGTCAGATTGCATACTGTCTCCTCCATGGCTACTCTCAAGGGTTTTGGCTGCG
GTGGATTGTTTTGTTTTCTCTAGAATCTAAAGAGGTTATCACAACGGCTTTGCAATTTGA
AAACTTTTCATGTTTGGGGAGATCAAAGATGGTTTTTTTTTATACCTTACTTGTAGAGA
GGATTGGAAGCAGCGAATAGCTGCAACCGGTCTGTTATGGATACTAATACATCTGGAGA
AGAATTATTAGCTAAGGCAAGAAAGCCATATACAATAACAAAGCAGCGAGAGCGATGGAC
TGAGGATGAGCATGAGAGGTTCTAGAAGCCTTGAGGCTTTATGGAAGAGCTTGGCAACG
AATTGAAGAACATATTGGGACAAAGACTGCTGTTTCAGATCAGAAGTCATGCACAAAAGTT

CTTCACAAAGTTGGAGAAAGAGGCTGAAGTTAAAGGCATCCCTGTTTGCCAAGCTTTGGA
CATAGAAATTCCGCTCTCTCGTCTAAACGAAAACCAATACTCCTTATCCTCGAAAACC
TGGGAACAACGGTACATCTTCTCTCAAGTATCATCAGCAAAAGATGCAAAACTTGTTC
ATCGGCTCTTCTTACAGTTGAATCAGGCGTTCTTGGATTGGAATAATGCCGTTCTC
TGAGAAAACATCAACTGGAAAAGAAAATCAAGATGAGAATTGCTCGGGTGTCTACTGT
GAACAAGTATCCCTTACCAACGAAAACAGGTAGTGGCGACATTGAAACAAGTAAGACCTC
AACTGTGGACAACGCGGTTCAAGATGTTCCCAAGAAGAACAAGACAAAGATGGTAACGA
TGGTACTACTGTGCACAGCATGCAAACTACCTTGGCATTTCACGCAGATATTGTGAA
CGGGAATATAGCAAAATGCCCTCAAAATCATCCCTCAGGTATGGTATCTCAAGACTTCAT
GTTTCATCCTATGAGAGAAGAACTCACGGGCACGCAAACTTCAAGCTACAACAGCATC
TGCTACTACTACAGCTTCTCATCAAGCGTTTCCAGCTTGTCTTACAGGATGATTACCG
TTCGTTTCTCCAGATATCATCTACTTTCTCCAATCTTATTATGTCAACTCTCTACAGAA
TCCTGCAAGTATGCTGCAAGTACATTGCTGCTTCCGCTGCTGGCTTATGCGAGTGTGCG
GAATTCTGGTGATTTCATCAACCCCAATGAGCTCTTCTCCTCCAAGTATAACTGCCATTGC
CGCTGCTACAGTAGCTGCTGCAACTGCTTGGTGGGCTTCTCATGGACTTCTTCTGTATG
CGCTCAGCTCCAAATAACATGTTTCCATTCTCAACTGTTGCACTTCCAAGTCCAGCAAT
GACTGAAATGGATACCGTTGAAAATACTCAACCGTTTGAGAAAACAAACACAGCTCTGCA
AGATCAAACTTGGCTTCGAAATCTCCAGCTTCATCATCTGATGATTGAGATGAGACTGG
AGTAACCAAGCTAAATGCCGACTCAAAAACCAATGATGATAAAATTGAGGAGGTTGTTGT
TACTGCCGCTGTGCATGACTCAAACTGCCCAGAAGAAAATCTTGTGGACCGCTCATC
GTGTGGCTCAAATACACCTTCAGGGAGTGACGCAGAACTGATGCATTAGATAAAATGGA
GAAAGATAAAGAGGATGTGAAGGAGACAGATGAGAATCAGCCAGATGTTATTGAGTTAAA
TAACCGTAAGATTAAATGAGAGACAACAACAGCAACAACAAATGCAACTACTGATTCTGTG
GAAGGAAGTCTCCGAAGAGGGTCGTATAGCGTTTCAGGCTCTCTTGTCAAGAGAAAGATT
GCCTCAAAGCTTTTCGCTCCTCAAGTGGCAGAGAATGTGAATAGAAAACAAAGTGACAC
GTCAATGCCATTGGCTCCTAAITTCAAAAGCCAGGATTCTTGTGCTGCAGACCAAGAAGG
AGTAGTAATGATCGTGTGGAACATGCAAGAGTCTTAAACGAGACAGACAGGATTTAA
GCCATACAAGAGATGTTCAATGGAAGTGAAAGAGAGCCAAGTTGGGAACATAAAACATCA
AAGTGATGAAAAGTCTGCAAAAGGCTTCGATTGGAAGGAGAAGCTTCTACATGACAGAC
TTGGAGGTAAATAAATAACATCCACATTTTATCAATATCTTTAAATCTAGTGTAGTAG
TTTGCTTCTCCAATCTTATGAAAGAGACTTTTAATTTTCCCTCCGAACATTCTTTGGT
CATGTCAGGTTCTGTACCATATTACCCATGTCTTGTCTCTTGTCTGTGTTGTGTATGC
TACTTGTGGTCTATATGTCATCTGCTACTACTGTTAATTAACCATTAAGCAATGGATTG
TCTTTA

>G680 Amino Acid Sequence (domain in AA coordinates: 24-70)
MDINTSGEELLAKARKPYTITKQRRWTEDEHERFLEALRLYGRWQRIEBHIGTKTAVQ
IRSHAQKFFTKLEKAEVKGIPVCQALDIEIPPPRPKRKPNTPYPRKPGNNGTSSSQVSS
AKDAKLVSASSSQLNQAFDLLEKMPFSEKSTGKENQDENCSESVSTVNKYPLPTKQVSG
DIETSKTSTVDNAVQDVPKKNKDKGNDGTTVHSMQNYPWPHFADIVNGNIAKCPQNHPS
GMVSQDFMFHPMRBETHGHANLQATTASATTTASHQAFPAHSDQDYRSFLQISSTFSNL
IMSTLLQNPAHAHAATFAASVWPYASVGNSSSTPMSSSPSITAIAAATVAAATAWWA
SHGLLPVCPAPITCVPFSTVAVPTPAMTEMDTVENTQPFQNTALQDQTLASKSPASS
SDDSDETGVTKLNADSKTNDKIEEVVVTAHVHDSNTAQKKNLVDRSSCGSNTPSGSDAE
TDALDKMEKDKEDVKETDENQPDVIELNNRKIKMRDNNNNATTDSWKEVSEEGRIAFQ
ALFARERLPQSFSPPQVAENVNRKQSDTSMPLAPNFKSQDSCAADQEGVVMIGVGTCKSL
KTRQTGFKPKYKRCSMEVKESQVGNINNQSDEKVKRLRLEGEAST*

>G867 (64..1098)

CACAAACACAAACACATTTCTGTTTTCTCCATTGTTTTCAAACCATAAAAAAAACACAGAT
TAAATGGAATCGAGTAGCGTTGATGAGAGTACTACAAGTACAGGTTCCATCTGTGAAACC
CCGGCGATAACTCCGGCGAAAAAGTCGTCGGTAGGTAACCTTATACAGGATGGGAAGCGGA
TCAAGCGTTGTGTTAGATTGAGAGAACGGCGTAGAAGCTGAATCTAGGAAGCTTCCGTCG
TCAAAATACAAAGTGTGGTCCACAACCAACGGAAGATGGGGAGCTCAGATTTACGAG
AAACACCAGCGGTGTGGCTCGGGACATTCAACGAAGAAGACGAAGCCGCTCGTGCCTAC
GACGTCGGGTTTACAGGTTCCGTCGCGTGACGCCGTACAAAATTTCAAAGACGTGAAG
ATGGACGAAGACGAGGTGATTTCTTGAATTTCTATTGAAATCTGAGATCGTTGATATG
TTGAGGAAACATACTTATAACGAAGAGTTAGAGCAGAGTAAACGGCGTCGTAATGGTAAC
GGAAACATGACTAGGACGTTGTTAACGTCGGGTTGAGTAATGATGGTGTCTTCTACGACG

GGGTTTAGATCGGCGGAGGCACTGTTTGAGAAAGCGGTAACGCCAAGCGACGTTGGGAAG
CTAAACCGTTTGGTTATACCGAAACATCAGCGAGAGAAACATTTCCGTTACCGTCAAGT
AACGTTTCCGTGAAAGGAGTGTGTTGAACTTTGAGGACGTTAACGGGAAAGTGTGGAGG
TTCCGTTACTCGTATTGGAACAGTAGTCAGAGTTATGTTTTGACTAAAGGTTGGAGCAGG
TTCGTTAAGGAGAAGAATCTACGTGCTGGTGACGTGGTTAGTTTCAGTAGATCTAACGGT
CAGGATCAACAGTTGTACATTGGGTGGAAGTCGAGATCCGGGTCAGATTTAGATGCGGGT
CGGGTTTGGAGATTGTTCCGAGTTAACATTTACCGGAGAGTTCAAGAAACGACGTCGTA
GGAAACAAAAGAGTGAACGATACTGAGATGTTATCGTTGGTGTGTAGCAAGAAGCAACGC
ATCTTTACGCCTCGTAACAACCTCTTCTTTTTTTTTTCTTTGTTGTTTAAATAATTT
TAAAAAATCCATTTTCGTTTTCTTTATTTGCATCGGTTTCTTTCTTCTGTTTACCAA
GGTTCATGAGTTGTTTGTGTTGATTGATGAACGTAAATTTTATTTATAGGATAAATTT
TAAAAAAAAAAAAAAAAAAAAA

>G867 Amino Acid Sequence (domain in AA coordinates: 59-124)
MESSVDESTTSGSICETPAITPAKKSSVGNLYRMGSGSSVLDSENGVEAESRKLPS
KYKGVVPQPNGRWGAQIYEKHQRVWLGTFFNEEDEAARAYDVAVHRFRRRDAVTNFKDVKM
DEDEVDFLNSHSEIIVMLRKHTYNEELEQSKRRRNNGNMTRTLTLTSLGSLNDGVSTTG
FRSAEALFEKAVTPSDVGKLNRLVIPKHAEKHFPLPSSNVSVKGVLLNFEDVNGKVWRF
RYSYWNSSQSIVLTGWSRFVKEKNLRAGDVVSFRSNGQDQQLYIGWKSRSGLDLAGR
VLRLFGVNISPESRRNDVVGKRVNDTEMLSLVCSKKQRIHAS*

>G956 (1..840)
ATGGAGGAGACAGAAAAGAATAAGGGCAGCATAAGTATGGTTGAGGCTAATCTACCTCCT
GGTTTGTAGATTCCATCCTAGAGACGACGAGCTCGTCTGTGACTACTTAATGAGAAGAACC
GTTTCGAGCCTCTATCAACCAGTTGTCTTGATCGACGTCGATCTTAACAAATGCGAGCCT
TGGGACATTCTCAAACGGCGAGAGTGGGAGGGAAGAATGGTACTTTTACAGCCAAAAA
GACCGTAAATACGCAACAGGCTACAGAACAACCGGGCTACGGCCACCGGTTATTGGAAA
GCCACCGGGAAGATAGAGCAATCCAAAGAAACGGTGGTCTTGTGGGTATGAGAAAGACA
CTTGTTGTTTACCGAGGTCGATCCCCTAAAGGTCGTAAAACTGATTGGGTCTATGCATGAG
TTTCGTCTCCAAGGAAAACCTTCTCACTCCCTAATTTCTCTCGAGGAAGAGTGGGTA
TTGTGTAGAGTTTCCACAAGAAGCAGCAACGGAGCTGATATAGACGACATCACAAGGAGC
TGCTCTGATGCAACAGCTTCTGCATTCATGGACTCTTACATCAACTTCGACCATCATCAC
ATCATCAATCAGATGTACCTGCTTCTCCAATAATTTGTACATAACCAAACCAACCAA
TCCGGTTAATCTCCAAGAATCCAGCCCATGTTTAAATGCTTCCCCTGATCAAATGATT
CTCAGAACTTTGCTAAGTCAACTCACAAAAAAGTCGAAGAATCACAGAGTCGTGGAGAC
GGAAGCTCAGAGACCAATTGACCGACATTGGCATCCCAAGCCATGCATGGAATTACTGA

>G956 Amino Acid Sequence (domain in AA coordinates: TBD)
MEETEKNKGSISMVEANLPPGFRFHPRDELVDYLMRRTVRSLYQPVLIDVDLNLKCEP
WDIPQATARVGGKEWYFYSQKDRKYATGYRTNRATATGYWKATGKDRAIQRNGLVGMKRT
LVFYRGRSPKGRKTDWVMHEFRLQGLLHHSFNSLEEEWVLCRVFHKNSNGADIDITRS
CSDATASAFMDSYINFDDHHIINQHVPCFSNNLSHNQTNQSGLISKNSPLFNASPDQMI
LRTLLSQLTKKVEESQSRGDGSSESQSLTDIGIPSHAWNY*

>G996 (53..1063)
CGATCGATCTTGAATTGATTCTTTGTAGTATTTTATTACATATATATATAGATGGGAAG
ACATTATGTTGTTACAAACAGAACTGAGGAAAGGACTTTGGTCTCCTGAAGAAGATGA
GAAGCTTCTTCGTTACATCACTAAGTATGGTCATGGTTGCTGGAGCTCTGTCCCTAAACA
AGCTGGTTTACAGAGATGTGAAAAAGTTGTAGATTAAAGATGGATAAATTATTTAAGACC
AGATTTGAAGAGAGGAGCATTTTCTCAAGATGAAGAAATCTCATATTGAACITCATGC
CGTCTTGGCAATAGATGGTCTCAGATAGCTGCACAGCTTCTTGGGAAGAACCGACAATGA
AATCAAGAATCTTTGGAATTTCTGTTTGAAGAAGAAATGAGGCTGAGAGGAATTGACCC
GGTTACACACAAGCTCTTAACCGAAATCGAAACCGGTACAGATGACAAAACAAACCGGT
TGAGAAGAGTCAACAGACCTACCTCGTTGAGACTGATGGCTCCTCTAGTACCACTACTTG
TAGTACTAACCAAAACAACAACACTGATCATCTTTATACCGGAAATTTTCGGTTTTCACG
GTTAAGTCTAGAAAACGGTTCAAGAATCGCAGCCGTTCTGACCTCGGTATCTGGATTCC
CCAAACCGGAAGAAACCATCATCATATGTCGATGAACCATCCCTAGTGCAGTGGTACT
ACCCGGTTCAATGTTCTCATCCGGTTTAAACCGGTTATAGATCCTCCAATCTCGGTTAAT
TGAATTGGAAAACTCAATCTCAACCGGGCCAATGATGACAGAGCATCAGCAAATTCAGA
GAGTAACTACAACAATTCAACATTCTTTGGAATGGGAATCTGAATTGGGGATTAACAAT
GGAGGAAAATCAAAATCCATTCAACATATCGAATCATTCAAATTCGTCCTTATACAGTGA

TATAAAATCAGAGACCAATTTTTTTGGCACAGAGGCTACAAATGTTGGTATGTGGCCATG
TAACCAGCTTCAGCCTCAGCAACATGCATATGGCCATATATAAATCTTCTGTATATTAT
AA

>G996 Amino Acid Sequence (domain in AA coordinates: 14-114)
MGRHSCCYKQKLRKGLWSPEEDEKLLRYITKYGHGCWSSVPKQAGLQRCGKSCRLRWINY
LRPDLKRGAFSQDEENLIIEHVLGNRWSQIAAQLPGRDNEIKNLWNSCLKKKLRLRG
IDPVTHKLLTEIETGTDDKTKPVEKSQQTYLVETDSSSTTCTSTNQNNNDHLYTGNFG
FQRLSLENGSRIAAGSDLGWIPIQTGRNHHHHVDETIPIAVVLPGSMFSSGLTGYRSSNL
GLIELENSFSTGPMTEHQIQESNYNNSTFFGNGNLNWGLTMEENQNPFITISNHSNSSL
YSDIKSETNFFGTEATNVGMWPCNQLQPQQHAYGHI*

>G1946 (90..1547)
TCTCACCTATTGTAAAAATCACCAGTTTTCGTATATAAAACCTAATTTTCTCAAAATTCC
CAAAATATTGACTTGGAAATCAAAAATCCGAATGGATGTGAGCAAAGTAACCACAAGCGACG
GCGGAGGAGATTCAATGGAGACTAAGCCATCTCCTCAACCTCAGCCTGCGGCGATTCTAA
GTTCAAACGCGCCTCCTCCGTTTCTGAGCAAGACCTATGATATGGTTGATGATCACAATA
CAGATTTCGATTGTCTCTTGGAGTGCTAATAACACAGTTTTATCGTTTTGGAAACCACCGG
AGTTTCGCTCGCGATCTTCTTCTTAAGAACTTTAAGCATAATAATTTCTCCAGCTTCGTTA
GACAGCTTAATACCTATGGTTTTCAGGAAGGTTGACCCAGATAGATGGGAATTTGCGAATG
AAGGTTTTTTAAGAGGTGAGAAGCACTTGCTACAATCAATAACTAGGCGAAAACTGCCC
ATGGACAGGGACAGGGACATCAGCGATCTCAGCACTCGAATGGACAGAACTCATCTGTTA
GCGCATGTGTTGAAGTTGGCAAATTTGGTCTCGAAGAAGAAGTTGAAAGGCTTAAAAGAG
ATAAGAACGTCCTTATGCAAGAACTCGTCAGATTAAGACAGCAGCAACAGTCCACTGATA
ACCAACTTCAAACGATGGTTTCAGCGTCTCCAGGGCATGGAGAATCGGCAACAACAATTAA
TGTCATTCTTGCAAGGCAGTACAAAGCCCTCATTTTCTATCTCAATTCTTACAGCAGC
AGAATCAGCAAAACGAGAGTAATAGGCGCATCAGTGATACCAGTAAGAAGCGGAGATTCA
AGCGAGACGGCATTGTCGTAATAATGATTCTGCTACTCTGATGGACAGATAGTGAAGT
ATCAACCTCCAATGCAAGGAGCAAGCAAGCAATGTTTAAACAGCTTATGAAGATGGAAC
CTTACAAAACCGCGCATGATGGTTTCTTCTAGGTATGGTACGCTACTACCGAGGGAA
CAGAGATGGAGACTTCATCAAACCAAGTATCGGGTATAACTCTTAAGGAAATGCCTACAG
CTTCTGAGATACAGTCATCATACCAATTGAAACAACCTCCTGAAAATGTTTCGGCAGCAT
CAGAAGCAACCGAGAACTGTATTCTTCACTGATGATCTAATCTTCCCAGCTTCACTC
ATATGCTACCGGAAAATAATTAGAGAAGCCTCCAGAGAGTTTCATGGAACCAACCTGG
GAGGTTCTAGTCCATTACTAGATCCAGATCTGTTGATCGATGATTCTTGTCTTCGACA
TTGACGACTTTCCAATGGATTCTGATATAGACCTGTTGATTACGGTTTACTCGAACGCT
TACTCATGTCAAGCCCGGTTCCAGATAATATGGATTCAACACCAGTGGACAATGAACAG
AGCAGGAACAAAATGGATGGGACAAAATAAGCATATGATAATCTGACTCAACAGATGG
GTCTCTCTCTCTGAAACCTTAGATCTCTCAAGGCAAAATCCTTGATTTTGGGAGTTTT
TAAAGTCTTTGAGGTACACAGTCCCTGAGAGCAGCATATTCAT

>G1946 Amino Acid Sequence (domain in AA coordinates: 32-130)
MDVSKVTTSDGGGDSMETKPSQPQPAAILSSNAPPPFLSKTYDMVDDHNTDSIVWSAN
NNSFIVWKPPPEFARDLLPKNFKHNNFSSFVRQLNTYGFVKVDPDRWEFANEGFLRGQKHL
LQSIITRRKPAHQGGQGHQHSQHSNGQNSSVSACVEVGKFGLEEEVERLKRDKNVLMQELV
RLRQQQQSTDNQLQTMVQRLQGMENRQQQLMSFLAKAVQSPHFLSQFLQQNQQNESNRR
ISDTSKKRRFKRDGIVRNND SATPDGQIVKYQPPMHEQAKAMFKQLMKMEPYKTGDDGFL
LGNGTSTTEGTEMETSSNQVSGITLKEMPTASEIQSSSPIETTPENVSAASEATENCIPS
PDDLTLPDFTHMLPENNSEKPPESFMEPNLGGSSPLLDPLLIDDSLSFDIDDFPMDSDI
DPVDYGLLERLLMSSPVPDNMDSTPVDNETEQEQNGWDKTKHMDNLTQQMGLLSPETLDL
SRQNP*

>G217 (84..2618)
cttcgttcttaccgagttccacgagcattagcttcagagaccttgaattggagtgcggtt
ggatcaaaaacagttgagcgaagatgaggattatgattaagggaggtgttggagaaca
ccgaagatgagattctcaaagccgctgatgaagtatggttaagaaccaatgggctcgga
tctcgtctctctcgttcgtaagtctgctaaacagtgtaaagctcgctggtacgagtggt
tcgatccatctatcaaaaagactgaatggaccagagaagaagatgagaagcttctacatc
ttgctaaactctgcctactcaatggagaactattgctcctatttggtgggtcgtaaccat
ctcaatgtcttgaggtatgagaagctccttgatgcagcatgcactaaggatgaaaatt
atgatgcagcggtgatccacgaaaattacgtcctggtgagattgatccgaaccagaag

caaaagcctgctcgctcctgatccggtagacatggacgaagatgagaaagaaatgctttctg
aagcaagagctagattggctaacacgaggggaaagaaggctaaaagaaaagctagagaaa
aacaacttgaggaagctagaaggcttgccttctctgcaaaaaagaagagaactaaaagcag
ctgggattgatggaaggcataggaagaaagagaaaggaatcgactataatgcagaaa
ttccttttgaaaagagggcacctgcgggattttatgatactgcggatgaagatcgctcctg
ctgatcaagtaaaatttccaactaccattgaagaacttgaaggaaaaagaagagctgatg
tagaagcacatttacgcaacaagatggtgcaaggaataaaattgctcagagacaggatg
ctccagcagctatatgcaagcaacaagctgaatgatccggaagtgttaggaagaggt
caaagctgatgttaccaccaccgcagatttcagaccacgagctagaagaaattgctaaga
tgggctatgccagtgccttcttgcgagaatgaggagctaacagaaggcagtgctgcta
ctcgtgcacttttggcaaattactcacaacaccaaggcaaggaatgacacccatgagga
cacctcaaagaactcctgctggttaaagggtgatgctattatgatggaagcagaaaacctgg
ccagattaagagactctcagacacctttgctaggaggagaaaatcctgagttgcaccctt
ctgacttcactggggtcactccgagaaagaaggagattcaaacgcctaataccaatgttga
cccttcaatgactcctggtggtgctggtcttactccaagaattggcttgacgccatcaa
gggattgggtccttcttcttatgacacccaaagggaactcccttcagggtgaacttcaca
ttaacgaagacatggacatgcagcaaaagtgcaaaacttgagaggcagagacgagaggaag
ctagaaggagtttactcgtctggtttgactgggcttctcagccaaagaacgagtaccaa
tagttgcacaacctcctcctgaggaaaagtgaagagccagaagagaaaattgaggaagaca
tgtcagacaggatagcgaggggaaaaggcgaggagaagaagcaagacaacaggcattgctta
agaagagatccaaggtcttgtagagagatcttcttagacccccagctgcttcattggcag
taattaggaactcgttgctttcagctgatggagacaaaagtctgttgttctcctactc
cgattgaggttgtagataaaaatggtaagagaggagcttctacagttgctggagcatgata
atgcaaagtatccgcttgatgacaaagctgagaagaagaaggagccaagaacctgacca
accgttctgcttctcaagttcttgaattgacgattttgatgaaaatgagctccaaggagg
ctgacaaaatgataaaggaggaggggaaagtcttctgtgtgtgtcaatgggacatgagaaca
agacacttgatgattttgtagaagctcacaacacatgcgtgaatgatctcatgtatttcc
cactcgaagcgttacgagctctcaagttgctggtggaacgcggacaaaagttgcagctt
ttcaggaggagatggagaattgagaaaaaagatggaggaggatgagaagaaggcagaac
acatgaaggccaagtacaaaacttatacaaaagggtcatgagaggaggcagagaccgtgt
ggacccaaatagaggcgacattgaagcaggctgagattggtggaacagaagtagatgct
ttaaagcattgaagaggcaagaagagatggctgcattcttttaggaaaaagaatttgcaag
aggaagtgtataaagcaaaaggaaacagagagtaaaactgcagactcgctatgggaatatgt
tggcaatggttgaaaaagcagaggagataatggtcgggttccgagcacaggcattgaaga
aacaagaggatgttgaaattctcacaactgaaagaagctaagctagccactggagagg
aaggagacatagccatagccatggaagcttctgcataaaaacttgagttttgtattgctt
acaagttttaaggagacgtagcttgactttgtattggttaagtttttttaatatgagtcac
gactttgtaaaaagggttatgatataattctctgtttgtatgctttgcaagagtcaagaaat
ttgaatgcttcaggatcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa

>G217 Amino Acid Sequence (conserved domain in AA coordinates: 8-67)

MRIMIKGVWKNTEDEILKAAVMKYGKNQWARISSLLVRKSAKQCKARWYEWLDPSIKKT
EWTREDEKLLHLAKLLPTQWRTIAPIVGRTPSQCLERYEKLLDAACTKDENDYDAADDP
KLRPGEIDPNPEAKPARPDPVDMDEKEMLSEARARLANTRGKAKRKAREKQLEEARR
LASLQKRRELKAAGIDGRHRKRKRKGIDYNAEIPFEKRAPAGFYDTADEDRPADQVKFPT
TIEELEGRKRADVEAHLRKQDVARNKIAQRQDAPAILQANKLNDPEVVRKRSKMLMLPPP
QISDHELEEIAKMGYASDLLAENEELTEGSAATRALLANYSTPRQGMTMPMRTPQRTAG
KGDIMMEAEENLARLRSQTPLLGGENPELHPSDFTGVTPRKKEIQTPNPMLTPSMTPGG
AGLTPRIGLTPSRDSSSFMTPKGTFPRDELHINEDMDMQSAKLERQRREARRSLRSG
LTGLPQPKNEYQIVAQPPPEESEBPEEKIEEDMSDRIAREKAEFEARQALLKKRSKVLQ
RDLPRPPAASLAVIRNSLSADGKSSVVPPTPIEVADKMVREELLQLLEHDNAKYPLDD
KAEEKKGAKNRTNRSASQVLAIDDFDENELQEADKMIKEEGKFLCVSMGHENKTLDDFVE
AHNTCVMDLMYFPTRSAVELSSVAGNADKVAAFQEMENVRKKMEDEKKAHEMKAKYKT
YTKGHERRAETVWTQIEATLKQAEIGGTEVECFKALKRQEEMAASFRKKNLQEEVIKQKE
TESKLQTRYGNMLAMVEKAEEIMVGFRAQALKKQEDVEDSHKLKEAKLATGEEEDIAIAM
EASA*

>G2192 (92..2971)

CGGAAAGAGATCAACCAACGATAGAGGAGAAGAAGAACTTGCATACGCAAAAAAATTTC

CCGGGAAAATTCCAGAACTGCTTTGGAAAAATGTGCGAGCCCGATGATAATTCCGCTAG
AAACGGCGTCACTACTCAACCTTCGAGGTCAAGGAGCTTCTAATGGATGTTGACGACTT
AGATCTTGACGGTTCATGGCCACTAGATCAAATCCCTTACTTATCCTCATCGAATCGCAT
GATTTCTCCGATTTTGTCTCCTCTTCTCTGAGCAGCCTTGCTCGCCTCTCTGGGCTTT
CTCCGACGGTGGAGGAAATGGTTTTCCACCACGCAACCTCCGGTGGCGATGATGAGAAGAT
CAGCTCTGTCTCCGGTGTTCCTTCTTTCCTCTCGCCGAGTATCCTCTCTCCTCCCTTA
CTCTTCTCCATCAGCAGCTGAGAACACAACAGAGAAGCATAACAGTTTCCAGTTTCCGTC
TCCATTGATGAGCCTAGTCCCACCAGAGAACACAGACAACCTACTGTGTGATCAAAGAGAG
GATGACTCAGGCGCTTCGATACTTCAAAGAATCAACCGAACACACGTTTTGGCTCAGGT
CTGGGCTCCTGTGAGAAAGAATGGTCGTGATTTGCTGACGACTTTGGGTCAACCTTTTGT
TCTTAATCCTAATGGTAATGGGCTTAATCAATACAGGATGATCTCTCTCACATATATGTT
TTCTGTGATAGTGAAAGTGACGTAGAGCTCGGACTCCCGGGTTCGAGTTTTCCGTCAGAA
ATTGCCTGAATGGACTCCAAATGTTTCAGTACTATTCCAGCAAAGAATTCTCGCGGCTTGA
TCACGCCTTGCACTACAACGTGCGTGGTACACTGGCCTTGCTGTCTTAATCCCTCTGG
TCAGTCCGTCATAGGTGTGTGGAACTTATAATGACCTCAGAGAAGATTCACTATGCACC
CGAAGTGAGCAAAAGTTTGCAAAGCCCTTGAGGCGGTAAATCTGAAAAGCTCGGAAATACT
TGATCACCACAAACACAGATATGCAATGAGAGTCGCCAAAACGCGCTTGCTGAGATTCT
CGAAGTGCTGACAGTTGTATGTGAGACCCATAACTTGCCTCTCGCTCAGACTTGGGTTCC
ATGTACGATGGGAGCGTTCTTGCCAATGGTGGCGGTCTAAAGAAAACTGCACCAGCTT
TGACGGTAGCTGCATGGGTCAAATCTGCATGTCTACAACCGACATGGCCTGCTATGTCGT
GGATGCTCATGTCTGGGGCTTTAGAGATGCCTGTCTTGAACACCATCTCCAGAAAGGCCA
GGGAGTCGTGGACGAGCTTTTCTCAATGGTGGCTCATGTTTCTGCAGAGACATACCAA
GTTCTGCAAAACGCGAGTACCCACTAGTCCATTATGCGCTCATGTTCAAGTTGACCACTTG
TTTTGCAATATCTCTCCAGAGCTCTTACACGGGCGACGACAGTTACATTCTTGAATTTTT
TCTTCTCTCGAGTATAACAGACGACCAAGAGCAAGATTGCTGTTGGGTTCTATTTTGGT
GACAATGAAAGAACATTTTTCAGAGTCTGAGGGTTGCATCTGGGGTTGACTTTGGTGAAGA
TGACGACAAATTGTCTTTTCGAGATCATCCAAGCATTACCGGACAAGAAGGTTTCATTCAA
AATAGAATCCATTTCGAGTTCCTTTTCTGGTTTAAAGTCAAATGCAACAGAGACGATGTT
GATTCCTCAGCCTGTGGTTTCAGTCTTCTGATCCAGTAAATGAGAAAATCAACGTGGCCAC
TGTTAACGGTGTGGTTAAGGAGAAGAAGAAAACAGAGAAAAAGCGTGGGAAGACTGAGAA
AACAAATCAGTCTAGATGTACTTTCAGCAGTATTTCACTGGAAGTCTCAAAGACGCTGCAA
GAGCCTAGGAGTTTGGCCGACGACAATGAAGCGAATTTGCAGGCAACACCGGAATCTCGCG
GTGGCCATCGAGGAAGATCAAGAAAGTGAATCGTTCAATCACAAGCTGAAACGAGTCAT
CGAATCTGTTCAAGGTACTGATGGAGGCCTCGACCTGACTTCCATGGCCGTTAGTTCCAT
CCCTTGACACACGGTCAAACATCAGCACAGCCACTAACTCACCAATGGTTCCAAACC
ACCTGAGCTACCAAACACCAATAATTACCTAACCATTTGTCAGGTGATCAGAGTCCGAA
CGAGCCAAATGGTTTCGCTGAGTTACCACCAAGCAATGGTCAAGCGATCACGAACGGT
GGATGAGAGCGCTGGGACTCCAACCTCTCATGGCTCATGTGACGGTAACCAATTAGATGA
ACCGAAAGTACCCAAATCAAGATCCGCTCTTTCAGGGTTGGTGGATCACCCGGGCTCCTTTT
TCCACCTTATTCTAGAGATCATGATGATCTGCAGCTTCTTTCGCAATGCCGAACAGGCT
TCTTGGTTCTATAGACCATTTCGAGGAATGCTCATTGAAGACGCTGGAAGTTCAAAGA
TCTGAGAAATCTTGTCCCTACTGCAGCATTGACGATAAGTTTCAAGACACAACTGGAT
GAACAATGATAATAATAGCAACAACAACCTTATACGCTCCCCCAAAGGAAGAGGCCATTGC
AAATGTTGCATGCCGAACCATCAGGCTCAGAAATGAGAACGGTAACAATCAAAGCAAGTTA
CAAAGACGACATAATACGGTTCAGAATATCCTCGGGTTCAGGTATAATGGAATTGAAGGA
TGAAGTGGCTAAGAGGCTGAAAGTTGATGCAGGAACGTTTCGATATCAAGTATCTTGACGA
TGATAACGAAATGGGTTTAAATAGCTTGTGATGCTGATCTTCAAGAATGTCTCGAGATCCC
TAGATCCTCCCGCAEGAAAATCGTAAGGCTCTTAGTTCATGATGTAACGACAAATCTAGG
GAGCTCCTGCGAGAGCACTGGAGAATTGTGACCTGATAATTCATTCGAACCTCTTTGTAA
ATAG

>G2192 Amino Acid Sequence (conserved domain in AA coordinates:600-700)

MCEPDDNSARNGVTTQPSRSRELLMDVDDLDDGWSWPLDQIPYLSNNRMISPIFVSSSS
EQPCSPLWAFSDGGNGFHHATSGDDEKISSVSGVPSFRLAEYPLFLPYSSPSAENTT
EKHNSFQFPSPLMSLVPPENTDNYCVIKERMTQALRYFKESTEQHVLAQVWAPVRKNGRD
LLTTLGQPFVLNPNNGNLNQYRMISLTYMFSVDSSESDVELGLPGRVFRQKLPEWTPNVQY
YSSKEFSRLDHALHYNVRGTLALPVFNPSGQSCIGVVELIMTSEKIHYAPEVDKVCKALE
AVNLKSSEILDHQTQICNESRQNALAEILEVLTVCETHNLPLAQTWVPCQHGSVLANG

GGLKKNCTSFDFGSCMGQICMSTTDMACYVVDAAHVWGFDRDACLEHHLQKGQGVAGRAFLNG
GSCFCRDITKFKCTQYPLVHYALMFKLTTCTFAISLQSSYTGDDSYILEFFLPSSITDDQE
QDLLLGSILVTMKEHFQSLRVASGVDFGEDDDKLSFEIIQALPDKKVHSHKIESIRVPFSG
FKSNATETMLIPQPVVQSSDPVNEKINVATVNGVVKEKKKTEKKRGKTEKTISLDVLQOY
FTGSLKDAAKSLGVCPPTMKRICRQHGISRWPSRKIKKVNRSITKLKRVIESVQGTGGGL
DLTSMVSSIPWTHGQTSAPLNSPNGSKPPELPNTNNSPNHSSDHPNEPNSPELPP
SNGHKRSRTVDESAGTPTSHGSCDGNQLDEPKVPNQDPLFTVGGSPGLLFPYSRDHVS
AASFAMPNRLGSDHFRGMLIEDAGSSKDLRLNLCPTAAFDDKFQDTNWMNNDNNSNNNL
YAPPKEEAIANVACEPSGSEMRTVTIKASYKDDIIRFRISGSGIMELKDEVAKRLKVDA
GTFDIKYLDLDDNEWVLIACDADLQECLEIPRSSRTKIVRLLVHDVTTNLGSSCESTGEL*
>G504 (69..1040)

CGTCGACCTCTTGACGATCATGAGACTGATTTCTGTAATAATATCGTCATTATATCAAATT
AGAAGTTGATGGAAAACATGGGGGATTCGAGCATAGGGCCGGCCATCCGCATCTCCCTC
CCGGGTTTCGGTTTACCCGACTGATGAGGAACAGTAGTTTACCTCAAGAAGAAAG
CAGATTCTGTTCCACTTCCAGTCTCAATCATCGCAGAGATTGATCTTTACAAGTTTGATC
CTTGGGAGCTTCCAAGCAAGGCGAGTTTTGGAGAGCAGAGTGGTACTTCTTAGTCCTC
GGGATCGGAAGTATCCAATGGGGTAGGCCAAACCGGCAGCAACTTCCGGTTATTGGA
AAGCAACGGGAACCGATAAACCGATATTTACGTGCAATAGTCACAAGGTTGGTGTCAAGA
AAGCGCTTGTTTTTTACGGTGGAAAGCCTCCTAAAGGGATAAAAACAGATTGGATCATGC
ATGAATATCGCCTCACTGATGGTAACCTTAGCACTGCGGCTAAGCCGCTGACTTAACCA
CGACAAGGAAAAACCACTACGGCTAGACGATTGGGTTCTATGTAGGATCTATAAGAAGA
ATAGTTCAAAAGACCAACAAATGGAGAGAGTATTACTTAGAGAGGATCTAATGGAAGGCA
TGCTCTCAAAATCATCTGCTAATTCTTCTTCTACATCAGTACTAGACAACAACGACAACA
ATAATAACAATAACGAAGAACACTTTTTTCGACGGTATGGTCGTTTCTTCAGACAAACGTT
CCTTGTGTGGTCAATACCGAATGGGCCACGAGGCCTCAGGATCATCTTCATTCCGATCTT
TCTTATCGAGCAAGAGGTTTTCATCATACAGGTGATCTCAACAATGATACTACAATGTCT
CTTTTGTTCGATGCTTAGTGAGATTCTTCAGAGTTCGGGGTTTCATGCAAATGGTGTATA
TGGATACGACGCTCTCTCTAGCTGATCATGGGGTTTTAAGACAGGCGTTTCAGCTTCCTA
ACATGAAGTGGCACTCATAATCTATATAGATATATATGTGTGTATCATATATGTATCTAT
GCAGGCCTAATATAGTTTACACATAAATCATCTGGGGCGGCCGCT

>G504 Amino Acid Sequence (domain in AA coordinates: TBD)
MENMGDSSIGPGPHLPFGFRFHPTDEELVVHYLKKKADSVPLPVSIIEIDLYKFDPE
LPSKASFGHEHWYFFSPDRKYPNGVRPNRAATSGYWKATGTDKPIFTCNHSHKVGKAL
VFYGGKPPKGIKTDWIMHEYRLTDGNLSTAAKPPDLTTRKNSLRLLDDWVLCRIYKKNSS
QRPTMERVLLREDLMEGMLSKSSANSSSTSVLDNNDNNNNMNEEHFFDGMVVSSDKRSLC
GQYRMGHEASGSSSFGSFLSSKRFHHTGDLNNDNYNVSFVSMLEIPQSSGFHANGVMDT
TSSLADHGVLRQAFQLPNMNWHS*

>G622 (248..2620)
TCTTTCTTTCTCAATTCGCCGTCAAAATCTTCTCTTTCTCTTCCCCGCCGGTCTCTCA
CCAATCCTCTGATCTCTCTACACACGAACCTTTGATTTTGACCAACGTCGATGCATGTTT
ATGACTAGTCTCTCTCAATCCTTCAATTTATCAATTCACGTCGATTTTCGTATCCGAT
TCGTTGTTCTAGCTCTTTGTGTGGTGTAGGGTTTTAAGATTTTGAATTGGGGTTGGA
GTTGTGTATGTTTGAAGTCAAAATGGGGTCAAAGATGTGCATGAACGCTTCATGTGGTAC
GACTTCTACTGTGAATGAAGAAAGGTTGGCCTCTTCGATCTGGTCTTCTCGCTGATCT
CTGTTATCGTTGCGGATCTGCGTATGAGAGTTCTCTATCTGTGAACAATTTTCATAAGGA
CCAATCTGGTTGGAGGGAATGCTATTTGTGTAGCAAGAGACTACATGTGGATGCATTGC
TTCTAAGGTAACGATTGAGTTAATGGACTATGGTGGTGTGGTTGTAGTACATGTGCTTG
CTGCCATCAACTCAATTTGAACACAAGGGGTGAGAATCCAGGTGTTTTAGCAGATTGCC
AATGAAAACGTTAGCTGATAGGCAACATGTAAATGGCGAAAGCGAGGAAGAAACGAAGG
CGATCTCTTTTCTCAGCCACTAGTCTAGGGCGGAGATAAAAGGGAAGAGTTTCATGCCTCA
CCGTGGGTTTTGGTAAGCTAATGAGTCCAGAAAGTACAACCACCGGCATAGGCTGGATGC
TGCTGGGGAAATGCATGAATCATCACCTTTACAGCCATCTTTAAATATGGGTTTGGCTGT
GAATCCGTTTAGCCCATCTTTTGAACCGAGGCTGTGAGGGAATGAAACACATCAGTCC
TTCTCAGTCCAACATGGTCCATTGCTCTGCTTCTAATATACTGCAAAAGCCATCAAGACC
TGCTATTTCAACTCCTCTGTGGCTAGTAAATCCGCTCAGGCGCGGATTGGAAGGCCTCC
TGTCGAAGGGCGAGGGAGAGGCCACTTGCTTCCGCGGTATTGGCCAAATATACGGATAA
AGAGGTTACAGCAGATCTCTGGAATTTGAATTTGAACATTGTACCTCTCTTTGAGAAAAC

TCTTAGTGCCAGTGATGCTGGTTCGCTAGTTCCTTCCAAAAGCCTGTGCAGA
GGCATATTTTCCCTCCGATTAGTCAATCCGAAGGCATTCCCTTTGAAAATCCAAGATGTGAG
GGGTAGGGAGTGGACGTTCCAGTTCAGATATTTGGCCCAATAACAATAGTAGAATGTATGT
TTTAGAAGGTGCTACTCCATGCATACAGTCCATGATGCTACAGGCTGGTGATACAGTAAC
TTTCAGTCGGGTTGATCCTGGCGGAAAACCTAATCATGGGTTCCAGGAAGGCAGCTAATGC
TGGAGACATGCAGGGTTGTGGGCTCACCAACGGAACATCAACTGAGGACACATCATCGTC
TGGTGTAAACAGAAAACCCACCCTCCATAAATGGTTCCTCGTGTATTTCACTAATACCGAA
AGAGTTGAATGGTATGCCTGAGAATTTGAACAGTGAGACTAACCGGGGCAGGATAGGTGA
TGATCCTACACGAGTTAAAGAGAAGAAGAGAACTCGAACCATTTGGTGCAAAAAATAAGAG
ACTTCTTTTGCATAGTGAAGAATCTATGGAGCTGAGACTCACTTGGGAAGAAGCTCAGGA
CTTGCTTCGTCCTCTCTAGTGTAAAGCCTACCATCGTTGTTCATGAGGAGCAAGAAAT
TGAAGAAATATGACGAACCTCCTGTCTTTGGAAAGAGGACTATAGTCACTACAAAACCTTC
AGGTGAACAGGAACGATGGGCAACTTGGCAGGACTGCTCTAAATGGAGAAGGTTACCTGT
AGATGCTCTTCTTTCTTTTAAATGGACATGTATAGACAATGTTTGGGATGTGAGTAGGTG
TTCATGTTCTGCACCGGAGGAGTCTGAAGGAACTTGAGAATGTTCTTAAAGTAGGTAG
AGAGCACAAGAAGAGAAGAACTGGGGAAAGACAGGCAGCACAAAGTCAGCAAGAACCGTG
TGGTTTGGACGCACCTGGCGAGTGCAGCAGTCTTAGGAGACACAATAGGCGAGCCAGAGGT
AGCGACCACGACCAGACATCCAAGGCACAGGGCTGGATGCTCTTCGATCGTGTGCATTCA
GCCACCAAGTGGGAAAAGGTAGGCACAAGCCTACATGTGGCTGCACTGTGTGTAGCACCGT
GAAGAGAAGGTTCAAGACGCTTATGATGAGGAGGAAGAAGAAGCAGTTGGAGCGCGATGT
AACAGCAGCAGAAGATAAGAAGAAGAGACATGGAAGTGGCTGAGTCTGATAAGAGTAA
GGAGGAGAAGGAAGTGAACACAGCGAGAATAGACCTGAACAGTGTCCATACAATAAAGA
AGATGTTGAAGCTGTTGCGGTGGAGAAAGAAGAGAGTCAAAAAAGAGCAATAGGACAGTG
TTCGGGCGTGGTGGCTCAAGACGCCAGTGTATGTTTAGGAGTTACAGAGTTAGAAGGAGA
GGGTAAGAATGTTCTGTGAAGAGCCGAGAGTTTCAAGCTGATATGGAAA

>G622 Amino Acid Sequence (domain in AA coordinates: TBD)
MFEVKMGSKMCMNASCSTTSTVEWKKGWPLRSGLLADLCYRCGSAYESSLFCEQFHKDQS
GWRECYLCSKRLHCGCIASKVTIELMDYGGVGCSTCACCHQLNLNTRGENPGVFSRLPMK
TLADRQHVNGESGGRNEGDLFSQPLVMGGDKREEFMPHRGFGKLMSPESSTTGHRLDAAG
EMHESPLQPSLNMGLAVNPFSPSFATEAVEGMKHISSPSQSNMVHCSASNILQKPSRPAI
STPPVASKAQARIGRPPVEGRGRGHLLPRYWPKYTDKEVQOISGNLNLNIVPLFEKTL
ASDAGRIGRLVLPKACAEAYFPPIISQSEGIPLKIQDVRGREWTFQFRYWPNNSRMVLE
GVTPCIQSMMLQAGDVTVFSRVDPGGKLIMSRKAANAGDMQGCGLTNGTSTEDTSSSGV
TENPPSINGSSCISLIPKELNGMPENLNSETNNGRIGDDPTRVKEKKRTRTIGAKNKRLL
LHSEESMELRLTWEEAQDLLRSPSPVKPTIVVIEEQEIEEYDEPPVFGKRTIVTTKPSGE
QERWATCDDCSKWRRLPVDALLSFKWTCIDNVWDVSRSCSAPEESLKELENVLVKGREH
KKRRTGERQAAQSQQEPCGLDALASAAVLGDTIGEPEVATTTTRHPRHRAGCSCIVCIQPP
SGKGRHKPTCGCTVCSTVKRRFKTLMRRKKKQLERDVTAAEDKKKKDMELAESDKSKEE
KEVNTARIDLNSDPYNKEDVEAVAVEKEESRKRAIGQCSGVVAQDASDLVGVTELEGEGK
NVREEPRVSS*

>G778 (50..1249)

TCTCAATAACACAAAACCTTTTAAACTAGTAAAATACACAGATTTTAGGATGAGCCAATG
TGTTCCAAACGTGCATCGATGATACTCCGGCAGCAGCCACCACCACCGTCCGCTCCAC
CACAGCCGCAGACATCCCCATATTAGACTACGAGGTAGCCGAGCTGACGTGGGAGAACGG
GCAACTAGGCTTGACGGCTTAGGTCCACCGCAGTGACGGCTTCGTCGACCAAGTACTC
CACAGCGCGCGGTGGAACGTTGGAGTCGATAGTGACCAAGCTACTCGCCTCCCTAACCC
TAAGCCCCACGGATGAGCTCGTCCCGTGGTTCCATCATCGCTCCTCAGGGCCGCGATGGC
AATGGACGCGCTTGTCCTTGTCTCAACCTAGTACACGAGCAGCAGAGCAAGCCTGGTGG
CGTTGGCTCCACCCGGGTGGGGTCATGTAGCGATGGTCGTACCATGGGCGGTGGAAAACG
AGCAAGAGTGGCACCAGTGGAGCGGGCGGGAGTCAGCGGCTGACCATGGACACTTA
CGACGTAGGTTTACCTCAACATCAATGGGCTCGCACGATAACACAATCGACGATCATGA
CTCCGTCTGCCACAGCCGCCACAGATGGAGGACGAAGAAGAGAAGAAAGCCGGAGGAAA
ATCATCAGTTTCAACCAAGAGAAGCAGAGCTGCTGTATTACATAACCAATCCGAACGTAA
GAGGAGAGATAAAATCAATCAAGAGATGAAGACTTTGCAAAAACCTGGTTCCCAATTCCAG
CAAGACGGATAAAGCATCTATGTTGGATGAAGTGATAGAGTATTTGAAGCAACTTCAAGC
ACAAGTGAGCATGATGAGCAGAATGAATATGCCTTCTATGATGCTTCTATGGCCATGCA
GCAACAACAACAACTACAAATGTCTCTCATGTCCAATCCCATGGGTTTAGGGATGGGCAT

GGGGATGCCCCGTCTCGGTCTCCTCGACCTTAATTCTATGAACCGAGCTGCTGCAAGCGC
TCCTAATATCCATGCCAACATGATGCCAAACCCATTTTGGCCATGAATTGTCCATCGTG
GGATGCTTCTTCCAATGACTCTCGATTTTCACTCTCCTCTCATCCCCGATCCTATGTCTGC
CTTTCTTGATGCTCTACTCAGCCAACGACGATGGAAGCGTATAGCAGGATGGCTACATT
ATATCAGCAAATGCAACAACAACTTCTCTCCTTTCGATCCAAAATGATTATTACTCAA
ACACCTCTATATAGTTTACGTCTATATATGTGTAGTCACATACATATATATATATTC
CATCATAATTATTTATTTATATGTATAGGCTTCTCATGAATTATGATATTATACGTATTA
CGTAAAAAA

>G778 Amino Acid Sequence (domain in AA coordinates: 220-267)
MSQCVPNCHIDDTAAATTTVRSTTAADIPILDYEVAELTWENGQLGLHGLGPPRVTAASS
TKYSTGAGGTLESIVDQATRLPNPKPTDELVPWFHHRSSRAAMAMDALVPCSNLVHEQQS
KPGGVGSTRVGS DGR TMGGGKRARVAPEWSGGGSQRLTMDTYDVGFSTSTSMGSHDNTI
DDHDSVCHSRPQMEDEEEKKAGGKSSVSTKRSRAAAIHQSERKRRDKINQRMKTLQKLV
PNSSKTDKASMLDEVIEYLKQLQAQVSMMSRMNMPMSMLPMAMQQQQQLQMSLMSNPMGL
GMGMGMPLGLLLDLNSMNRAAAAPNIHANMMPNPFPMNCPSPWDASSNDSRFQSPFLIPD
PMSAFLACSTQPTTMEAYSRMATLYQQMQQLPPSPNPK*

>G791 (173..877)
TTTTCTTTGGGTGTTTCTTCCACCAACGGCAGAAATCGATTGCGGCTTAAATCTCCCCCTC
CTTTGATCTCTCTGATCGCCGCCGGGAACATTCAATTCCCGGGAGTTCAACAAAAAA
AAACTCTCCGTTTTTATTTTTCCCCCTTTTTCACCGGTGGAAGTTTCCGGAGATGGTGTC
ACCCGAAAACGCTAATTGGATTGTGACTTGATCGATGCTGATTACGGAAGTTTCACAAAT
CCAAGGTCCTGGTTTCTCTTGGCCTGTTTTCAGCAACCTATTGGTGTTTCTTCTAACTCCAG
TGCTGGAGTTGATGGCTCGGCTGGAACCTCAGAAGCTAGCAAAGAACCTGGATCCAAAAA
GAGGGGGAGATGTGAATCATCCTCTGCCACTAGCTCGAAAGCATGTAGAGAGAAGCAGCG
ACGGGACAGGTTGAATGACAAGTTTATGGAATTGGGTGCAATTTTGGAGCCTGGAATCC
TCCCAAAACAGACAAGGCTGCTATCTTGGTTGATGCTGTCCGCATGGTGACACAGCTACG
GGGCGAGGCCCAGAAGCTGAAGGACTCCAATTCAAGTCTTCAGGACAAAATCAAAGAGTT
AAAGACTGAGAAAAACGAGCTGCGAGATGAGAAACAGAGGCTGAAGACAGAGAAAGAAAA
GCTGGAGCAGCAGCTGAAAGCCATGAATGCTCCTCAACCAAGTTTTTCCCAGCCCCACC
TATGATGCCTACTGCTTTTGCTTTCAGCGCAAGGCCAAGCTCCTGGAACAAGATGGTGCC
AATCATCAGTTACCCAGGAGTTGCCATGTGGCAGTTCATGCCCTCCTGCTTCAGTCGATAC
TTCTCAGGATCATGTCTTCTGCTCCTCTGTTGCTTAATCAAGAAAAATCATCAACCGGTT
TGCTTCTTGCTTCCGCTTAAAAGAAAAGTCTCCATTTGTTTTGCTCTCCTCTCTTCTCG
GCTTCTTAGTCTTATCCTTTTGCTTTGTCGTGTTATCATCGTAACGTGTTATCTGTTGAA
CAATGATATGACATTGTAACTCCAATTGCTTCGCGCAATGTTATCTATTACATGTAAA
TTTAAGTAGAGTTTGGCAAAAAAA

>G791 Amino Acid Sequence (domain in AA coordinates: 75-143)
MVSPENANWICDLIDADYGSFTIQPGFSPVQPIGVSSNSSAGVDGSAGNSEASKEPG
SKKRGRCESSSATSSKAKREKQRRDLNDKFMELGAILEPGNPKTDKAILVDAVRMVT
QLRGEAQKLKDSNSSLQDKIKELKTEKNELRDEKQRLKTEKEKLEQQLKAMNAPQPSFFP
APPMPTAFASAQGQAPGNKMVPIISYPGVAMWQFMPPASVDTSQDHVLRPPVA*

>G861 (158..880)
CTTCTCTCTCCTCCTCCATCTCTTCTCTTACTCTCTCTTTAATCATCTCTCATTTCTGA
ATCTTGATCCATCAAAATCAATCCCGTTCTCGAAAGATCCATTAAAAATCAAAACCTAAGC
TCTCTCTCTTGCTTCTAGGGTTTTTTTGTTCGTTGTGATGGCGAGAGAAAAGATTCAGAT
CAGGAAGATCGACAACGCAACGGCGAGACAAGTGACGTTTTTCGAAACGAAGAAGAGGGCT
TTTCAAGAAAGCTGAAGAACTCTCCGTTCTCTGCGACGCCGATGTCGCTCTCATCATCTT
CTCTTCCACCGGAAAACCTGTTTCGAGTTCTGTAGCTCCAGCATGAAGGAAGTCTTAGAGAG
GCATAACTTGCAGTCAAAGAACTTGAGAAGCTTGATCAGCCATCTCTTGAGTTACAGCT
GGTTGAGAACAGTGATCAGCCCCGAATGAGTAAAGAAATTGCGGACAAGAGCCACCGACT
AAGGCAAATGAGAGGAGAGGAACCTCAAGGACTTGACATTGAAGAGCTTCAGCAGCTAGA
GAAGGCCCTTGAACTGGTTTTGACGCGTGTGATTGAAACAAAGAGTGACAAGATTATGAG
TGAGATCAGCGAACTTCAGAAAAAGGAATGCAATTGATGGATGAGAACAAGCGGTTGAG
GCAGCAAGGAACGCAACTAACGGAAGAGAACGAGCGACTTGGCATGCAATATGTAACAA
TGTGCATGCACACGGTGGTGCTGAATCGGAGAACGCTGCTGTGTACGAGGAAGGACAGTC
GTCGGAGTCTATTACTAACGCCGGAACCTCTACCGGAGCGCTGTTGACTCCGAGAGCTC
CGACACTTCCCTTAGGCTCGGCTTACCGTATGGTGGTTAGAGATGGAACAATTCAAAGAA

GTTGATGGAGTGAGGAGAGTAATGTAAATCTTTTTAACTCGGTAGTAACAAGAGACAATG
TCTAAGTAGTGAATTCTCAAATGTTTGTGTAAGTTTCTGCCTATGGAAGAGGCTTTCATT
TTTATGATTTTCACTATGTATGATCTCTTCTCACTGCATTTCTGGTTAGTAACGGCTTGT
CACCGATAAACTTTCTCGTTATGGAAAGTTAGAATAAAAAAAAAAAAAAAAAAAAAA
>G861 Amino Acid Sequence (domain in AA coordinates: 2-57)
MAREKIQIRKIDNATARQVTFKRRRLFKKAEELSVLCDADVALIIFSSTGKLFEFCSS
SMKEVLERHNLQSKNLEKLDQPSLELQLVENS DHARMSKEIADKSHRLRQMRGEELQGLD
IEELQOLEKALETGLTRVIETKSDKIMSEISELQKKGMQLMDENKRLRQQTQLTEENER
LGMQICNNVHAHGAESENAAVYEEGQSSESITNAGNSTGAPVDSESSDTSRLRLGLPYGG
*

>G938 (1..1755)
ATGATGATGTTTAAAGAGATGGGAATGTATGGAAACATGGATTCTTCTCTTCTCCACA
TCTCTCGATGTGTGCCATTACCACAAGCTGAACAAGAACCTGTAGTTGAAGATGTCGAC
TACACCGATGATGAGATGGATGTGGATGAGCTTGAGAAGAGGATGTGGAGAGACAAAATG
CCTTTGAAACGTCTCAAGGAGCAACAGAGTAAGTGTAAGAAGGCGTCGATGGTTCGAAA
CAGAGGCAGTCGCAAGAGCAAGCTAGGAGGAAGAAAATGTCTAGAGCCCAAGATGGGATC
TTGAAGTATATGTTGAAGATGATGGAAGTTGTAAAGCTCAAGGCTTTGTTTATGGTATT
ATTCTGAGAAGGGTAAGCCTGTGACTGGTGCTTCGGATAATTTGAGGGAATGGTGGAAA
GATAAGGTTAGGTTTATCGTAATGGTCCAGCTGCTATTGCTAAGTATCAGTCAGAGAAT
AATATTTCTGGAGGGAGTAATGATTGTAAACAGCTTGGTTGGTCCAACACCGCATACGCTT
CAGGAGCTTCAGGACACGACTCTTGGTTTCGCTTTTATCGGCTTTGATGCAACATTGTGAT
CCACCGCAGAGACGGTTTCTTTGGAGAAAGGAGTTTCTCCACCTTGGTGGCCTAATGGG
AATGAAGAGTGGTGGCCTCAGCTTGGTTTACCAAATGAGCAAGGTCCTCTCTTATAAG
AAGCCTCATGATTGTAAGAAAGCTTGGAAAGTCGGTGTTTAACTGCGGTGATCAAGCAT
ATGTCGCCGGATATTGCGAAGATCCGTAAGCTTGTGAGGCAATCAAAATGCTTGCAGGAT
AAGATGACGGCGAAAGAGAGTGCTACTTGGCTTGCCATTATTAACCAAGAAGAGGTTGTG
GCTCGGGAGCTTTATCCCGAGTCATGCCCTCTCTTCTTCTTCTTCATCATTAGGAAGC
GGTTCGCTTCTCATTAAATGATTGTAGCGAGTATGACGTTGAAGGTTTCGAGAAGGAACAA
CATGGTTTTCGATGTGGAAGAGCGGAAACCAGAGATAGTGATGATGCATCTCTAGCAAGC
TTTGGGGTTGCTAAAATGCAACATTTTCCCATAAAGGAGGAGGTCGCCACCACGGTAAAC
TTAGAGTTTACGAGAAAGAGGAAGCAGAACAAATGATATGAATGTTATGGTAATGGACAGA
TCAGCAGGTTTACCTTGTGAGAATGGTCAAGTGTCTCACAGCAAAATGAATCTTGGATT
CAAGACAGGAGTTCAAGGGACAACCACAGATGGTTTGTCCATATAGAGACAATCGTTTA
GCGTATGGAGCATCCAAGTTTCATATGGGTGGAATGAACTAGTAGTTCTCAGCAACCA
GTCCAACCGATCGACCTATCGGGCGTTGGAGTTCCGGAAAACGGGCAGAAGATGATCACC
GAGCTTATGGCCATGTACGACAGAAATGTCCAAAGCAACCAACGCCTCTACTTTGATG
GAAAACCAAAGCATGGTCATTGATGCAAAAGCAGCTCAGAATCAGCAGCTGAATTTCAAC
AGTGGCAATCAAATGTTTATGCAAAAGGACGAACAACGGGGTTAACAATCGGTTCCAG
ATGATGTTTGAATTCGACACCATTCAGATATGGCAGCATTGATTACAGAGATGATTGGCAA
ACCGGAGCAATGGAAGGAATGGGGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAAGATGTA
TCAATATGGTTCTGA

>G938 Amino Acid Sequence (domain in AA coordinates: 96-104)
MMMFNEMGMYGNDFFSSSTSLDVCPLPQAEQEPVVEDVDYTDDEMDVDELEKRMWRDKM
RLKRLKEQQSKCKEVDGSKQRQSQEQARRKKMSRAQDGLKYLKMMEVCKAQGFVYGI
IPEKGPVGTGASDNLREWWDKVRFRDRNGPAAIAKYQSENNISGGSNDCNSLVGPTPHTL
LELDQTTLGSLLSALMQHCDPPQRRFPLEKGVSPWPWPNGNEEWWPQLGLPNEQGPPPYK
KPHDLKKAWKVGVLTAVIKHMSPDIAKIRKLVRQSKCLQDKMTAKESATWLAIINQEEVV
ARELYPESCPPLSSSSSLGSGSLINDCSEYDVEGFEKEQHGFDEERKPEIVMMHPLAS
FGVAKMQHFPIKEEVATTNLEFTRKRKQNDNMVMVMDRSAGYTCENGQCPHSMNLGF
QDRSRDNRHQMVCYPYRDNRLAYGASKFHMGMKLVVPQQPVQPIDLSGVVPENGQKMIT
ELMAMYDRNVQSNQTPPTLMENQSMVIDAKAAQNQQLNFNSGNQMFMQQGTNNGVNNRFQ
MVFDSTPFDMAAFDYRDDWQTGAMEGMGKQQQQQQQQQDVSIWF*

>G965 (73..1956)
GATTCTCTGTGTATGTCTGAATCCTTACAGGATCCAAGAGCTTTGGAAAAAGATATAAT
GAATAACAAGATATGGGTTTAGCTACTACAACTTCTTCTATGTACAAGATTATCATCAT
CACCAAGGAATCTTTCTTCTCTAATGGATTCCACCGATCATCATCAACCACTCATCAG
GAGGAAGTAGATGAATCCGCCGTCGTCTCCGGTGCTCAAATCCGGTTTATGAAACCGCC

GGAAATGTTGTCTGAAATGTTTGCTTACCCTGGCGGAGGTGGCGGCGGTTCCGGTGGAGAG
ATTCTTGATCAGTCTACTAAACAGTTGCTAGAGCAACAAAACCGTCACAACAACAACAAAT
AACTCAACTCTTCATATGTTATTACCAAATCATCATCAAGGTTTTGCTTTACCCGACGAA
AACACTATGCAGCCGAGCAACAACAACACTTTACATGGCCATCTTCCTCCTCCGATCAT
CATCAAAACCGAGATATGATCGGAACCGTCCACGTGGAAGGAGGAAAGGGTTTGTCTTTA
TCTCTCTCATCTTCATTAGCCGAGCTAAAGCCGAGGAATATAGAAGCATTTATTGTGCA
GCCGTTGATGGAATCTCTCTTCTTCTAACGCATCCGCTCATCATCATCAATTCAATCAG
TTCAAGAAATCTTCTTCTTGAGAATTTCTTCTCAACATCATCACCATCAAGTTGTGGA
CATTTTGGTTCATCATCATCATCTCCCATGGCGGCTTCTTCATCCATGGAGGGATCTAC
ACGTTGAGGAATTCGAAATATACGAAACCGGCTCAAGAGTTGTTGGAAGAGTTTGTAGT
GTTGGAAGAGGACATTTCAAGAAGAACAACCTTAGTAGGAACAACCTAAACCTAATACT
ACCGGTGGAGAGGAGGCGGAGGGTCCCTCGTCATCGCCGGAACAGCTAATGATAGTCCT
CCTTTGTCTCCGGCTGATCGGATTGAACATCAAAGAAGAAAAGTCAAGCTACTATCTATG
CTTGAAGAGGTGGACCGACGGTACAACCACTACTGCGAACAATGCAAAATGGTAGTGAAC
TCATTCGACCAAGTAATGGGTTACGGCGCGGCGGTTCCGTACACGACATTAGCTCAAAAG
GCAATGTCTAGGCATTTCCGGTGTTTGAAAGACGCGGTAGCGGTTAGCTTAAACGCAGC
TGTGAGCTTCTAGGGGATAAAGAGGCGGCGAGGGCTGCATCCTCGGGGTTAACCAAGGG
GAAACGCCGCGATTGCGTTTGCTAGAGCAGAGTTTTCGTGAGCAACGAGCGTTTCATCAT
ATGGGTATGATGGAGCAAGAGGCATGGAGACCGCAACGTGGTTTGCCTGAACGCTCCGTT
AATATCCTTAGAGCTTGGCTATTTCGAGCATTTTCTTAATCCGTACCCAAGCGATGCTGAT
AAGCACCTCTTAGCAGCAGAGCTGGTTTATCCAGAAATCAGGTGTCAAATTGGTTCATA
AATGCTAGGTTTCGCTTATGGAAACCAATGGTGGAGAGAGATGTATCAACAAGAAGCAAAA
GAAAGAGAAGAAGCAGAAGAAGAAAATGAAAATCAACAACAACAAGAAGACAGCAACAA
ACAAACAACAACGACACGAAACCAACAACAATGAAAACAACCTTCACTGTCATAACCGCA
CAAACCTCCAACGACGATGACATCGACACATCACGAAAACGACTCTTCATTCTCTCTTCC
GTCGCCGCCGCTTCTCACGGCGGTTACGACGCGTTACCGTCGCCACGTGTCAGCAAGAC
GTCAGTGACTTCCACGTCGACGGAGATGGTGTGAACGTCATAAGATTCCGGGACCAACAG
ACTGGTGACGTGTCTCTTACGCTTGGTCTACGCCACTCTGGCAATATTCTTGATAAGAAC
ACTTCTTTCTCEGTTAGAGACTTTGGAGATTTTTAGTCTTCTTTGTTTCTCAATTTATTC
ATC

>G965 Amino Acid Sequence (domain in AA coordinates: 423-486)
MGLATTTSSMSQDYHHHQGLFSFSNGFHRSSSTTHQEEVDES AVVSGAQIPVYETAGMLS
EMFAYPGGGGGGSGGEILDQSTKQLLEQQNRHNNNNSTLHMLLPNHQGFATDENTMQ
PQQQHQFTWPSSSSDHHQNRDMIGTVHVEGGKGLSLSLSSSLAAAKAEYRSIYCAAVDG
TSSSSNASAHHHQFNQPKNLLLENSSSQH HHQVVGHFSSSSSPMAASSSIGGIYTLRN
SKYTKPAQELLEEFCSVGRGHFKKNKLSRNNSNPNTTGGGGGGSSSSAGTANDSPPLSP
ADRIEHQRRKVKLLSMLLEEVD RRYNHYCEQMOMVNSFDQVMGYAAVPYTTLAQKAMSR
HFRCLKDAVAVQLKRSCELLGDKEAAGAASSGLTKGETPRLRLLEQSLRQQRAPHHMM
EQEAWRPQRLPERSVNILRAWLFEHFLNPYPSPDADKHLARQTGLSRNQVSNWF INARV
RLWKPMVEEMYQQEAKEREEAEENENQQQRRQQQTNNNDTKPNNNNENFTVITAQTPT
TMTSTHHENDSSFLSSVAAASHGGSDAFTVATCQQDVSD FHDVGDGVNVIRFGTKQTGDV
SLTLGLRHSGNIPDKNTSFSVRDFGDF*

>G1143 (54..677)

AAATAAGAAATATAAACAACCTTTTGTCTGAAAAATTATCAAAGAAGAAGAAATAAATGGGTG
GAGGAAGCAGATTTCAAGAACAGTGAGGATGAGCCGTAGGAAACAAGTAACAAAAGAGA
AGGAAGAAGATGAAAACCTCAAATCTCAAATCTGAAGCAGAGAGACGTAGAAGAGAGA
AGCTTCATTGTGCGCTTATGGCTCTGCGATCTCATGTCCCCATTGTCACCAACATGACTA
AAGCAAGTATTGTTGAAGATGCGATTACTTACATAGGAGAGCTTCAAACAATGTTAAGA
ATCTCTTAGAGACATTTATGAAATGGAAGAGCTCCTCCTGAGATTGATGAAGAACAAA
CGGATCCAATGATAAAACCTGAAGTTGAAACTAGTGATCTTAACGAAGAGATGAAGAAAC
TCGGAATCGAGGAGAATGTGCAATTGTGTAAGATTGGGGAGAGGAAGTTTGGTTAAAGA
TCATAACAGAGAAGAGAGATGGGATCTTTACTAAATT CATGGAGGTTATGAGATTTCTCG
GATTCGAGATTATCGATATTAGTCTAACAACCTCAAATGGAGCAATTCTTATTAGTGCCT
CTGTTT CAGACACAGGAACCTCTGTGATGTTGAACAGACAAAAGATTTTCTTTTGGAGTTA
TGAGAAGCAATCCATAAGTATTAATTATATACATCTTGGAATTTCTTGATCTAATAACA
TTTCCATTGGTTTTTTATTAACATTGTTGTTCCATTTTAAATATGATATGATTCAGATGAAA
AAGAGTTTGTGTTACAAGCCAATGA

>G1143 Amino Acid Sequence (domain in AA coordinates:33-82)
MGGGSRFQEPVRMSRRKQVTKEKEEDENFKSPNLEAERRRRREKLHCLMALRSHVPIVTN
MTKASIVEDAITYIGELQNNVKNLLETFFHEMEEAPPEIDEEQTDPMIKPEVETSDLNEEM
KKLGIENVQLCKIGERKFWLKIITEKRDGIFTKFMEVMRFLGFEEIIDISLTTSSNGAILI
SASVQTQELCDVEQTKDFLLEVMRSNP*

>G1190 (209..2020)

TCCTGTCCCAAAACCAAAAGACTTGAGAGTGTGTCTTTAGAGAGAGATCTTCTCTCTTTT
ATCTTAGGACTCTCACTTCTTATCTCAAATCTACTTCAACTCTATTTCCAGTCTCCACAT
TTTCCCAACAATTTCAACTCTTGTCTCTTCTCCAAAGTAAAAACAATCGTTGCAAG
TGAGGTTTGGTTTTGGTGTATAGAATTATGAAGAGCGGGAAGCAATCTTCGCAACCTGA
AAAGGGTACTTCCAGGACTCTTGTCACTGACTGTCTGTTTATCGCATTGTGCGGTTTCTC
CTTCTACCTCGGTGTATATTTTGTCTGAGAGAGACAAGATTGTAGCCAAGGATGTCAC
AAGGACGACTACAAAGGCTGTAGCTTCCCCTAAAGAACCTACAGTACTCCTATTCAAAT
CAAATCCGTTTCTTCCCAGGAGTGGGTCAGAGTTCCAAGATTACACCCCGTGACCCGA
TCCAAAGAGGTGGAGAAGTATGGTGTCCATCGCTTAAGTTTCTTGGAGCGTCATTGTCC
TCCGGTATATGAAAAGATGAGTGTGATTCCACCACCAGACGGGTATAAACCGCCTAT
AAGATGGCCCAAGAGCCGAGAACAGTGTGTTGATACAGGAACGTGCCATTATGATTGGATCAA
TAAGCAAAAGTCTAACCAGCATTGGCTTAAGAAAGAAGGAGATAAGTTCCATTTCCTTGG
TGGTGGTACCATGTTCCCTCGTGGAGTTAGTCACTATGTTGATTGATGCAAGATCTGAT
TCCTGAAATGAAAGACGGAACAGTCAGGACCGCCATTGATACTGGCTGTGGGGTTGCGAG
CTGGGGAGGCGATCTTTTGGACCGTGGGATACTATCACTCTCTCTGCTCCAAGAGATAA
CCATGAAGCTCAGGTTCAATTTGCTCTTGAACGTGGAATTCCTGCGATTCTCGGGATCAT
CTCTACGCAACGTCTCCCTTTTCTTCAAATGCATTTGATATGGCTCATTGTTCAAGATG
TCTTATTCCTGGACAGAATTTGGTGGAACTTATTTACTTGAGATTACCGTATAGTTTCG
ACCTGGAGGTTTGGTGTCTTTTCTGTTCCACCTGTGAACTATAATAGACGATGGCGTGG
ATGGAACACAACCATGGAAGATCAGAAATCTGACTACAACAAGCTTCAGTCACTTCTAAC
CTCCATGTGTTTCAAAAAGTACGCTCAAAAAGATGACATAGCCGTGTGCGAGAACTCTC
AGACAAATCTTGCTATGACAAAATCGCTAAGAACATGGAAGCTTACCTCCCAAATGTGA
CGACAGTATAGAACCTGATCTGCTTGGTACACTCCACTCCGTCCTTGGCGTGGTTGCCCC
GACACCTAAAGTCAAGAAGTCTGGTCTCGGATCAATCCCAAAATGGCCCGAGAGGTTACA
TGTCGCGCCCGAGAGAATCGGTGATGTTTACGGAGGGAGTGCGAACAGTTTGAACACGA
TGATGGTAAATGGGAAGAACAGAGTTAAGCATTACAAGAAAGTTTTACCAGCTCTTGGGAC
AGACAAGATAAGAAATGTTATGGATATGAACACTGTTTATGGAGGTTTCTCTGCGGCCCT
CATTGAGGATCCCATTGGGTTCATGAACGTTGTATCATCGTACAGCGCAAATTCGCTTCC
TGTTGTCTTTGATCGCGGTCTCATCGGGACTTACCACGACTGGTGCAGAGCTTTCTCAAC
GTATCCAAGACATATGATCTTCTTACCTCGACAGTCTTTTACCTTGGAGAGTCACAG
GTGTGAGATGAAGTACATTTTGCTAGAGATGGACAGGATCTTGGGCCGAGTGGATATGT
TATAATCCGAGAATCGAGTTATTTTATGGACGCAATCACACGTTAGCGAAAGGGATAAG
GTGGAGTTGCCGAGAGAGAGAGACTGAGTATGCAGTCAAAGTGAGAAGATTCTGGTTTG
CCAGAAAAGCTATGGTTTTTCTGTCAAACCAACCTCTTGTATGAGACCACCTGTATCATAG
TGTTTATCATCTCTGTGATGCACACTACAGAGAGAAGGATCTAGTCCTTTGAGTCCAAG
ATATAGCTCTATAACAATCTCCTTTTTTTTGTCTCTTTAATTTCTTGGGTATTTACCGG
TATAGATTGATATTATATATTTTTTAATTATATTTTTAATATATAGATATATAGTATGT
GGTTTAAACACTATTATTATCAAGGTCTTAAAGATTGCTTTGCAAGAGTTAAAAAATGT
TGGAGTAAGGACCTCTTGATTAATAAATTGACTGACGCGAGCAAA

>G1190 Amino Acid Sequence (domain in AA coordinates: entire protein)

MKSGKQSSQPEKGTSRILSLTVLFIACGFSFYLGIFCSERDKIVAKDVTRTTTKAVAS
PKEPTATPIQIKSVSFPEGSEFQDYTPCTDPKRWKYGVHRLSFLERHCPPVYEKNECL
IPPPDGYKPIRWPKSREQWYRNVYPYDWINKQSNQHWLKEGDKFHPFGGGMFPFRGV
SHYVDLMQDLIPEMKDGTVRTAIDTGCGVASWGGDLLDRGILSLSLAPRDNHEAQVQFAL
ERGIPAILGIIISTQRLPFPNSAFDMAHCSRCLIPWTEFGGIYLLIHRIVRPGGFVWLSG
PPVYNRRWRGWNNTMEDQKSDYNKLQSLTSMCFKKAQKDDIAVWQKLSDKSCYDKIA
KNMEAYPPKDDSIEPDSAWYTLRPPCVVAPTPKVKSGLSIPKWPERLHVAPERIGDV
HGGANSLSKHDDGKWKNRVKHYKKVLPALGTDKIRNVMDMNTVYGGFSAALIEDPIWVMN
VVSSYSANSLPVVDRGLIGTYHDWCEAFSTYPRTYDLLHLSLFTLESHRCMKYILLE
MDRILRPSGYVIIRESSYFMDAITTLAKGIRWSCRRETEYAVKSEKILVCQKKLWFSSN
QTS*

>G1198 (230..1675)

TCTTTTCAAATTTCCAATCATTTGATCAACTAATCAAGAATTAATTATAAGACTTTGCAAT
CTCTCTCCCTCTCCCTCTCCCTAGCTAGTCTCTCTTGTGTTTCTTAACCTCGAGCTTCTC
TCAATAGTGATTATCATCTTTTTCATCATTTCAAGATTTAATGTGTTTTGCAGAAAAGAG
ACTAATCAAGAAGAGATATCATCAATTGAAGCTGTTTTCTTGAGTAGAGATGGCGAACCA
TAGAATGAGCGAAGCTTCAAAACCATAACCACAATCATCATCTTCCTTATTCATTATTCA
TGGTCTCAACAACAATCATCCATCTTCTGTTTTATTAAACCAAGATGGATCGTCCAGTTT
CGATTTTGGAGAGCTAGAAGAAGCAATTGTTCTGCAAGGTGTCAAGTATAGGAACGAGGA
AGCCAAGCCACCTTTATTAGGAGGAGGAGGAGGAGCTACGACTCTGGAGATGTTCCCTTC
GTGGCCAATCAGAATCACCAAACTCTTCTACTGAGAGTTCCAAGTCAGGAGGAGAGAG
CAGCGATTCAAGGATCGGCTAATTTCTCCGGCAAAGCTGAAAGTCAACAACCGGAGTCTCC
TATGAGTAGCAACATCATCTCATGCTTCAACCTCATCATAATAACATGGCAAACCTCAAG
TTCAACATCTGGAGCTTCCCTTCCACTTCTCGAACTTTAGCTCTCTTAAACCTTCGGAAGA
TAAGAGGAAGGCTACAACCTTCAGGCAAACAGCTTGATGCTAAGACGTTGAGACGTTTGGC
CCAAAATAGAGAAGCTGCTCGCAAAGCCGTCTTAGGAAAAAGGCGTATGTGCAACAGCT
AGAATCAAGTAGGATAAAGCTTTCCCAATTGGAGCAAGAACTTCAGCGAGCTCGTTCTCA
GGGGCTGTTTATGCGGTGTTTGTGGACCAACAGGACCTAACATCACTTCCGGAGCTGCAAT
ATTTGACATGGAATATGGGAGATGGCTAGAGGATGATAACCGGCATATGTCCGAGATTTCG
AACCAGCTTTCAGGCTCATTTATCTGACAATGATTTAAGGTTGATCGTTGACGGTTACAT
TGCTCATTTTGTAGAGATATTCGGATTAAAGCCGTGGCAGCGAAAGCCGATGTTTTTCA
CCTCATCTATTGGGACATGGATGTCCCGAGCCGAACGTTGTTTTATTGGATGGCTGGTTT
CCGTCCATCCGACCTAATCAAGATATTGGTGTGCGCAATGGATCTATTGACGGAGCAACA
ACTGATGGGAATATATAGCCTACAACACTCGTCGCAACAAGCAGAGGAGGCTCTCTCGCA
AGGCCTCGAACAACCTTCAGCAATCTCTCATCGATACTCTCGCCGATCTCCAGTCATTGA
CGGAATGCAACAATGGCTGTGCGCTCTCGGAAAGATCTCTAATCTCGAAGGCTTTATCCG
CCAGGCTGATAACTTGAGGCAGCAGACCGTTCCAGCTGAGGCGGATCTTGACCGTCCG
ACAAGCTGCACGGTGTTCCTAGTCATCGGAGAGTACTATGGACGGCTCAGAGCTCTTAG
CTCCCTTTGGTTGTACGCCCCACGAGAGACACTGATGAGTGATGAAACCTCTGTCAAAC
GACGACGGATTTGCAGATTGTTTCAGTCATCTCGGAACCACTTCTCCAATTTCTGAATGGA
ATGAAACTTTGTATAACTAAAGGCCAAGTTTCATTGTCTGTCTGTAATTTTACCTATTTT
CTTTAAAGTTGTACTAGAGAAAAGATAGGATCTTCTCTCG

>G1198 Amino Acid Sequence (domain in AA coordinates: 173-223)

MANHRMSEATNNHNNHLPYSLIHGLNNHNPSSGFINQDSSSFDFGELEEAIVLQGVKY
RNEEAKPPLLGGGGGATTTLEMFPSWPIRTHQTLPTSSKSGGESSDSGSANFSGKAESQQ
PESPMSSKHHLMLQPHNNMANSSSTSGLPSTSRTLAPPKPSEDKRKATTSKGQLDAKTL
RRLAQNREARKSLRKKAYVQQLSSRIKLSQLEQELQRARSQGLFMGGCGPPGPNITS
GAAIFDMEYGRWLEDDNRHMSEIRTGLQAHLSNDLRLIVDGYIAHFDEIFRLKAVAACA
DVFHLLIGTWMSPAERCFIWMAGFRPSDLIKILVSQMDLLTEQQLMGIYSLQHSSQQAEE
ALSQGLEQLQSLIDLTAASPVIDGMQMAVALGKISNLEGFIRQADNLRQOTVHQLRRI
LTVRQAARCFVLIGBEYYGRLRALSSLWLRSRPRETLMSDETSCQTTTDLQIVQSSRNHFSN
F*

>G1226 (212..1159)

CTGCATTTATTAAAGACAGTTTAGAAAGTGTCACCCCTAAAGGAATGTTTTAGTTTAG
AGGAAAGAGAGAGAAGAAGAGCAGCAGCAGAAGTTGTTAATTTGAAGACTATTTGAGGA
AAGACACCTATATCTAAATACTCAAAGTTACAAAATATTAATTTCAGAAAACAGTTCCAT
TAGAGAGACTCATAAAGCTTCTCATCTAATTATGAGTGGATTGATGAGTTTTGGTGAATT
AGAAGACCAATTTGGTCAGATTTTCAGACACTACTATGGAAGAGAAGATACCATTTCTGCA
AATGCTTCAATGCATAGAACACCCCTTTTACAACAACAGAACCAATCAGTTTCTCCAATC
ACTTCTCCAGATCCAAACCTTAGAATCAAAGAGCTGTCTCACCTTGAAACAAACATCAA
AAGAGATCCGGGTCAAACAGATGACCCGAAAAGGATCCAAGAACAGAAAACGGAGCAGT
AACGGTCAAAGAAAAAGAAAACGGAAACGTACAAGAGCTCCAAGAACAAGACGAAGT
TGAAAACCAAAGGATGACTCACATTGCCGTGCAACGTAATCGAAGACGACAAATGAACGA
ACACTTAACTCTCTCCGATCTCTCATGCCCTCCTTCGTTTCTTCAACGGGGTGACCAAGC
TTCGATTGTAGGAGGGGCAATAGATTTTCATCAAGGAAGTACAGCAACTCTTGCAATCTCT
AGAAGCTGAGAAAAGGAAGGATGGAACGTGATGAAACTCTTAAACGGCGTCTGTGTTCTC
ATCTTCGTCTCTTGATGCACTAACTCTTCTATTTCTAGCGTGTCTACGACGTCGGAAAA
TGGATTTACGGCGAGATTCGGCGGTGGAGATACGACAGAAGTGGAGGCTACGGTGATACA

GAACCATGTGAGCTTAAAAGTTCGGTGTAAGAGAGGAAAAACGACAGATCTTAAAAGCTAT
TGTCCTCGATTGAAGAACTAAAGCTTGGGATTCTACATCTCACTATCTCTTCTTCTTGA
CTTTGTCTACTCTCTTCAATCTCAAGATGGAAGATGGTTGTAAATTAGGATCAGCAGA
TGAGATAGCGACAGCCGTTTCATCAGATCTTCGAGCAAATCAACGGTGAAGTCATGTGGTC
AAATCTTAGTCGAACTTAGTTGACTTTTGACTCCTAGTAACGTGTGTAACTTTAGGTTA
CAAAGAAAAGGACGTGATATAAATAAGAAAAACCAAGAGGTGAAATTTTGGGAGTTTT
AATTATTATCTTATACTTTTGGATTTTAGATTAGTAGCAAACCTCGCAGTGTTCTACGAT
GACATTATTATTGGTCACATGAAGGTTTAGGTTAAAAAAA
>G1226 Amino Acid Sequence (domain in AA coordinates:115-174)
MSGLMSEGELEDQFGQISDITMEEKIPFLQMLQCIHPFTTTEPNQFLQSLQLQITLESK
SCLTLETNIKRDGPQTDDEKPRTEGAVTVKEKRRKRTRAPKNKDEVENQRMTHIAV
ERNRRRQMNELHNSLRSLMPPSFLQRGDQASIVGGAIDFIKELEQLQLSLEAEKRDGTD
ETPKTASCSSSSSLACTNSSISSVSTTSENGFTARFGGDTTEVEATVIQNHVSLKVRCK
RGKRQILKAIVSIEELKLAILHLTISSSFDFVIYSFNLKMEDGCKLGSADIELATAVHQIF
EQINGEVMWSNLSRT*
>G1451 (124..2559)
TTTGTACTTCCGGAGCTAAAGAGTTATAGCTACTGTAGTAGCTGGAAGTGAAGAAGATTT
TTTAATAGATTGTACGGAATAATTAGGGTTTTCAAAGTTTGGTTTCTTGAAGTTGAATTA
GACATGAAGCTGTCAACATCTGGATTGGGTCAACAGGGTCATGAAGGAGAGAAGTGTCTG
AATTCTGAGCTATGCGATGCTTGTGCTGGACCATTAGTCTCTCTTCCATCATCTGGTAGT
CGAGTTGTTTACTTTCCACAGGGTCACAGTGAACAGGTAGCTGCTACAATAAGGAA
GTTGATGGTCACATACCCAATTACCAAGCCTACCACCACAATTGATATGCCAGCTCCAT
AATGTTACAATGCATGCAGATGTTGAGACGGATGAAGTCTATGCTCAAATGACACTTCAA
CCATTGACACCGGAGGAGCAGAAGGAAACATTTGTACCGATTGAGTTGGGGATACCGAGT
AAGCAACCTAGTAATTATTTTGTAAAGACTCTCACAGCTAGTGATACCAAGTACACATGGA
GGGTTTTCTGTTTCTTAGACGTGCTGCTGAGAAAGTGTTCCTCCATTGGATTACACACTG
CAGCCACCAGCTCAAGAAGTATTGCAAGGGATCTCCATGATGTTGAATGGAAGTTTAGG
CATATCTTTCCGGGACAGCCCAACGGCATCTCCTAACTACTGGATGGAGTGTCTTGTGTC
AGTGCCAAAGCGACTAGTAGCTGGAGATCTGTGCTATTTTCATCAGGAATGAAAAGATCAA
CTCTTTTTTGGGAATTCGTCATGCCACTCGGCCGCGAGACTATTGTACCATCATCTGTTTTA
TCTAGTGATAGCATGCATATTGGACTCCTTGCTGCTGCTGCACATGCTTCTGCAACTAAT
AGCTGTTTTCACTGTTTTCTTTTCATCCAAGGGCTAGCCAATCTGAGTTTGTGATACAATT
TCCAAGTACATTAAAGCCGTTTTTTCACACGCGTATTTTCAGTTGGGATGCGCTTTCGCATG
CTCTTCGAGACAGAAGAGTTCGAGTGTCCGAGGTACATGGGTACTATAACTGGTATTAGT
GATCTAGATTCTGTTTGGTGGCCAACTCTCATTGGCGATCTGTGAAGGTTGGTTGGGAT
GAATCGACTGCAGGGGAGAGACAGCCAAAGGGTTTTCTTATGGGAGATTGAGCCTCTGACT
ACCTTTCTCTATGTATCCATCTCTTTTTCTCTCAGACTAAAACGTCCATGGCATGCTGGC
ACATCATCTTTGCTGATGGAAGGGGTGATTGGGAAGTGGTCTAACATGGCTAAGAGGG
GGAGGTGGAGAGCAGCAAGGTTTGGCTTCTCTAAATTATCCATCTGTGGTTTTGTTTCCA
TGGATGCAACAAAGGCTGATCTCAGTCAAATGGGGACTGATAAATCAGCAATACCAA
GCAATGTTAGCTGCTGGGTTGCAGAACATCGGCGGTGGAGATCCTTTAAGACAGCAGTTT
GTACAGCTGCAAGAGCCTCACCACCAATATCTTCAACAATCAGCTTCCCATATTCTGAT
TTGATGCTTCAGCAGCAACAGCAGCAACAGCGTCACGCCATCTCATGCATGCTCAAACA
CAGATTATGAGTGAGAATCTTCCGACAGCAATATGCGACAAGAAGTTAGTAACCAACCA
GCTGGACAGCAGCAACAGCTACAGCAACCGGACCAAAATGCATATCTTAATGCTTTCAA
ATGCAAAATGGCCATCTTCAACAGTGGCAGCAGCAATCAGAGATGCCATCTCCCTCGTTC
ATGAAGTCAGATTTTACTGACTCAAGCAACAAATTTGCAACAAC'TGCTAGTCCGGCTTCT
GGAGATGGCAATCTTTTGAATTTTTCTATAACCGGTCACTCTGTACTCCCTGAGCAGTTA
ACAACAGAGGGGTGGTCTCCAAAAGCATCCAACACTTTTTCTGAACCGTTGTCACTTCCA
CAAGCCTATCTGGGAAGAGTCTTGCTCTAGAACCCGGAATCCGAGAAATCCCTCTCTT
TTCGGTGTGATCCCGACTCTGGACTCTTCTCCCGAGTACGGTTCCCGCTTTGCTTCT
TCATCAGGAGATGCTGAAGCTTCCCCTATGTCACTAACAGATTGAGATTTCAGAAATCC
TTATATAGCTGCATGCAAGACACAACATCATGAGTTATTGCATGGAGCTGGACAGATTAACT
TCGTCCAACCAACCAAGAACTTTGTAAAGGTTTATAAATCTGGTTTGGGTCGTTTCA
TTAGACATCTCCCGATTGAGCAGCTACACGAGCTGCGAGAAGAGTTAGGGAAGATGTTT
GCTATCGAAGGGTGTGTTGGAAGACCCCTTAGATCAGGCTGGCAGCTTGTATTGCTTGAC
AAGGAAAATGATATCTTCTCTTGGTGATGACCCATGGGAGTCATTTGTGAATAACGTT

TGGTACATAAAGATACTATCACCAGAAGATGTGCATCAAATGGGAGATCATGGAGAAGGC
AGTGGTGGGTATTCCCGCAAAACCCGACCCATCTCTAGAAGCTGCTTCGGTGTAGTCT
CATCATGCTACAACGCGGGAGCCCTTTGTTTCCCATTGGAAGTCGTTTCCACTCATCTTT
ATATGCCATTTCGTTCCGATCTCTCTCGTTTTGACGTTTTTAGAAAGAAACATAATCATAT
TTGTGAGTATGGGTCCTGAAACTTTAGGACGTACTTTAGCTTGTATTAGACAGACACTCT
CGTCATAAACATAAGAACCCTTTATGTAGCTGTCTCAGGGTAACTAACTTTTCTAG
>G1451 Amino Acid Sequence (domain in AA coordinates: 22-357)
MKLSTSGLGQOGHEGEKCLNSELWHACAGPLVSLPSSGSRVVYFPQGHSEQVAATTNKEV
DGHIPNYPSPQLICQLHNVTMHADVETDEVYAQMTLQPLTPPEQKETFFVPIELGIPSK
QPSNYFCKTLTASDTSTHGGFSVPRRAAEKVFPPLDYLQPPAQELIARDLHDVIEWKFRH
IFRGQPKRHLLTGWVSFVSAKRLVAGDSVIFIRNEKNQLFLGIRHATRPQTIVPSSVLS
SDSMHIGLLAAAAHASATNSCFTVFFHPRASQSEFVIQLSKYIKAVFHTRISVGMRFRL
FETEESVRRYMGTTTGISDLSVRWPNHWSVKVWGDESTAGERQPRVSLWEIEPLTT
FPMYPSLPLRLKRPWHAGTSSLPDGRGDLGSGLTWLRGGGGEQQGLPLNYPVGLFPW
MQQRDLDSQMGTDNNQQYQAMLAAGLQNIIGGGDPLRQOFVQLQEPHQQYLQSSASHNSDL
MLQQQQQQQASRHLMAQTQIMSENLPQONMRQEVSNQPAQQQQQLQPPDQAYLNAFKM
QNGHLQWQQQSEMPSPSFMKSDFTDSSNKFATTASPGDGNLLNFSITGQSVLPEQLT
TEGWSPKASNTFSEPLSLPQAYPGKSLALEPGNPQNPSLFGVDPDSGLFLPSTVPRFASS
SGDAEASPMSLTDSGFQNSLYSCMQDTTHELLHGAGQINSSNQTKNFVKVYKSGSVGRSL
DISRFSSYHELREELGKMFAIEGLLEDPLRSGWQLVFVDKENDILLGGDDPWESFVNNVW
YIKILSPEDVHQMDHGEHSGGLFPQNPTH*
>G1478 (1..354)
ATGTGTAGAGGGTTTGAGAAAGAAGAAGAGAGAAGAAGCGACAATGGAGGATGCCAAAGA
CTATGCACGGAGAGTCAAAAGCTCCGGTAAGCTGTGAGCTTTGCGGCGAGAACGCCACC
GTGTATTGTGAGGCAGACGCAGCTTTCTTTGTAGGAAATGCGATCGATGGGTCCATTCT
GCTAATTTTCTAGCTCGGAGACATCTCCGGCGCGTGATCTGCACGACCTGTCCGAAGCTA
ACTCGTCGATGCTTGTCTGGTGATAATTTTAAATGTTGTTTTACCGGAGATAAGGATGATA
GCAAGGATTGAAGAATAGTAGTGATCACAAAATTCCTTTGTGTTTCTCTGA
>G1478 Amino Acid Sequence (domain in aa coordinates: 32-76)
MCRGFEKEEERRSDNGGQRLCTESHKAPVSCBLCGENATVYCEADAFLCRKCDRWVHS
ANFLARRHLRRLVICTTCKRLTRRLVGDNFNVVLPEIRMIARIEBHSSDHKIPFVFL*
>G1496 (116..1123)
AAACCCACCAATAACTCAGAGCTTTTTTGCATTTTTTCCCATTCTCTATTTTGTGTTTGT
ACTTTTGGTCTCACTTTAAAGATCATAAGTTGAAAGATTCTGCAGAGAACAATATGTT
GGAAGGTCTTGTCTCTCAAGAAAGCTTGTCTCTTAAACTCTATGGACATGTCTGTACTTGA
AAGGCTTAAATGGGTACAACAGCAACAACAGCAACTGCAACAAGTTGTGTCCCATAGCAG
TAATAATTCACCTGAACCTTCTCAGATACTTCAGTTCCATGGAAGCAACAATGATGAGTT
GTTGGAGAGTAGTTTTACGCCAATTTCAAATGCTTGGATCTGGTTTTTGGACCAAACTATAA
CATGGGTTTTTGGTCTCCACATGAATCCATTTCAAGAAACAAGTAGCTGCCATATGGAACC
TGTGGATACAATGGAGGTTTTGTTGAAGACCGGTGAAGAAACCAGAGCCGTTGCCTTGAA
GAACAAGAGAAAACCAGAGGTTAAGACAAGGGAAGAGCAAAAGACAGAGAAGAAGATCAA
AGTAGAGGCTGAGACAGAGTCAAGCATGAAAGGAAAATCAAACATGGGAAACACTGAAGC
ATCTTCAGACACTTCAAAGGAGACATCGAAAGGAGCTTCAGAGAATCAGAAATTAGATTA
TATCCACGTGAGAGCTCGTCGAGGCCAAGCCACTGACAGACACAGCTTAGCAGAAAGGGC
GAGAAGAGAAAAGATCAGCAAGAAAATGAAATATCTGCAAGATATTGTGCCTGGATGCAA
TAAGGTCACAGGAAAAGCTGGTATGCTTGATGAGATCATCAATTATGTTCAATGTCTCCA
AAGACAAGTCGAGTTCTGTGATGAACTTGCTGTCTTGAACCCGGAAGCTAGAGCTTGC
CGTGGAAAGATGTATCCGTAAAAACAGGCTTACTTTACAAATGTAGTTGCTTCAAAGCAATC
AATAATGGTTGATGTGCCATTGTTTTCCGTTAGACCAGCAAGGATCTCTAGATTTGTCTGC
GATAAACCCGAACCAACGACATCTATCGAAGCTCCATCTGGAAGCTGGGAAACTCAATC
ACAGAGTCTCTACAACACATCTAGCCTCGGTTTTTCATTACTAAGCAAGATTCTTGAAC
AACATGGTTGACATCAATCAATCATCAAAATCAGAAGCAAATCTATTACATTTGCTCAT
CAAAGTAGTAATTTGCAAATTTGGTTAATGCATTATCCTTTGATCCTTTGTTTTCTGATAT
TTAAACCAGAAGAACTGGAGATAGCAATCCAATGATCTTGTCACCA
>G1496 Amino Acid Sequence (domain in AA coordinates: 184-248)
MLEGLVSQESLSLNSMDMSVLERLKWVQQQQQLQVSHSSNNSPELLQILQFHGSNND
ELLESSFQFQMLGSGFGPNYMGFGPPHESISRTSSCHMEPVDMEVLLKTGEETRAVA

LKNKRKPEVKTREEQKTEKKIKVEAETESSMKGKSNMGNTEASSDTSKETSKGASENQKL
DYIHVRARRGQATDRHSLAERARREKISKMKYLQDIVPGCNKVTGKAGMLDEIINYVQC
LQRQVEFLSMKLAVLNPELELAVEDVSVKQAYFTNVVASKQSIMVDVPLFPLDQQGSLLDL
SAINPNQTTTIEAPSGSWETQSQSLYNTSSLGFHY*

>G1526 (1..3090)

ATGGGAACGAAAGTCTCAGACGATCTTGTTCACCGTCAGATCAGTCGTGGGTTCGGAT
TACTCAGATATGGATATAATCAGGGCTTTACACATGGCGAATCATGATCCAACGGCTGCT
ATCAATATAATCTTCGACACTCCAAGTTTCGCCAAACCTGATGTAGCCACTCTACCCCG
AGCGGCTCTAATGGAGGGAAGCGAGTTGATAGTGGATTAAAGGGCTGTACTTTTGGTGAC
AGCGGAAGTGTTCGAGCGAATCATCGCGTGGAGGAAGAAAATGAGAGTGTAAATGGTGGA
GGAGAAGAGAGTGTTCAGGGAATGAGTGGTGGTTTGTGGTTGTTCTGAATTGGCTGGG
TTATCGACATGTAAAGGAAGGAAATTGAAGTCTGGTGATGAATTGGTGTTCACGTTTCCG
CATAGTAAAGGATTAAAGCCTGAGACTACGCCCTGGGAAGCGCGGTTTGGGCGGGGAAGG
CCAGCTTTGCGTGGTGCTTCTGATATCGTTAGGTTCTCTACAAAGGATTTCAGGAGAGATT
GGTAGAATACCAAACGAGTGGGCTCGGTGTCTTCTACCACTTGTGAGAGACAAGAAAATT
AGGATAGAAGGCAGTTGCAAGTCCGCCCTGAAGCTTTGAGCATCATGGATACAATTCTT
CTGTCTGTAAGCGTGTACATTAATAGTTCATGTTTCAAAGCATAGTGCAGCTTCATTT
AAGACAGCTAGTAATACGGCAGAGGAATCAATGTTCCATCCTCTCCCAAATCTCTTCGG
TTACTCGGTTTGTATCCCTTTAAGAAGGCAGAGTTTACTCCAGAGGATTTTACTCTAAG
AAGCGACCTTTGAGTTCCAAGGATGGTTCTGCTATTCTACTTCGTTGCTTCAATTAAAC
AAGGTCAAGAATATGAATCAAGATGCAAACGGAGATGAAAATGAGCAGTGTATCAGCGAT
GGTGATCTTGATAACATTGTTGGTGTGGGGACAGTTCTGGATTAAAGGAAATGGAAACT
CCACATACACTTCTGTGTGAGCTTCGTCCATACCAAAGCAGGCACCTTCATTGGATGACC
CAACTGGAGAAAGGAAATTGCACTGTATGAGGCAGCAACAATGCTTCACCCGTGTTGGGAA
GCATCTGTTTAGCAGACAAGAGGGAAGTGGTTGTCTACCTGAATCTTTTACTGTGTAT
GCTACAATACACTTCCCTAGCACACTTCAAATGGCAAGAGGAGGAATATTAGCAGACGCA
ATGGGTCTTGGAAGAGACTGTAATGACCATATCCCTTTTGCTTGCCCATTTCTTGGAAGCT
GCATCAACTGGGTTTCTATGCCCAACTATGAAGGAGACAAAGTGTATCAGCAGTTCTGTA
GATGATCTCACTAGTCCCCCGGTGAAGGCAACCAATTCTTAGGCTTTGATAAGAGGCTT
CTTGAACAAAAAAGTGTACTTCAAATGGTGGTAACCTGATTGTATGTCCGATGACACTT
TTAGGACAGTGGAAAGACAGAGATTGAAATGCATGCAAAGCCTGGGTCTCTATCTGTCTAT
GTTCACTATGGGCAAAGCAGGCCGAAGGATGCAAACTTCTTTCCAGAGTGTATGGTA
ATCACCACATATGGAGTTCTAACATCCGAATTCTCGCAAGAGAACTCAGCAGACCATGAA
GGAATTTATGCAAGTTTCGATGGTTTAGGATTGTTCTTGACGAGGCACATACCATCAAAAC
TCAAAAAGCCAAATTTCTTGCTGCTGCAGCTCTGGTTGCTGATAGGCGTTGGTGTCTT
ACGGGTACTCCTATTAGAACAACTCTGGAGGATTTATACAGCCTTCTACGGTTTTTGAGG
ATTGAACCATGGGGAACCTTGGGCATGGTGAATAAACTTGTCCAAAAGCCATTTGAAGAG
GGTGATGAGAGAGGGTTAAAGCTAGTGCAGTCTATCTTAAACCTATCATGCTTAGGAGA
ACAAAGTCTAGCACAGACCGAGAAGGAAGGCCGATTCTTGTCTACCCCTGCTGATGCA
CGGGTCATTTACTGTGAACCTTCGGAGTCTGAGAGGGATTCTACGACGCGCTATTTAAA
AGATCCAAGGTCAAATTTGATCAATTTGTTGAACAAGGCAAAGTCTTTCATAACTATGCT
TCGATCCTGGAACTGCTTTTGCCTCTTCGACAATGTTGTGATCACCCATTTTATGTAATG
AGTCGAGGGGATACAGCGGAATACTCTGATCTGAATAAGCTTCTTAAACGTTTCCTTAGT
GGAAAGTCTTCTGGCTTAGAAAGGGAAGGAAAAGATGTACCGTCAGAGGCTTTTGTTCAG
GAGGTGGTAGAGGAACTGCGCAAAGGAGAGCAAGGAGAGTGTCCAATATGCCTTGAAGCA
CTTGAGGATGCTGTATTAACGCCATGTGCTCATAGATTATGTCGTGAGTGTCTCTTGGCA
AGTTGGAGAAATTCTACTTCTGGGTTATGTCTGTGTGTAGGAACACTGTAAGCAAACAA
GAACTCATCACAGCACCAACCGAAAGTAGATTCCAGGTTGACGTGGAAAAGAATTGGGTG
GAATCATCGAAAATCACTGCTTCTGGAAGAGCTTGAAGGCTTTCGTTCTTCAGGCTCT
AAGAGCATTCTCTTAGCCAGTGGACCGCTTTCCTCGATCTCTCCAAATCCCCTCTCT
CGGAATAACTTTTCATTTGTCCCTCTTGATGGCAGCTAAGTCAGCAGCAACGAGAGAAG
GTCCTTAAAGAATTTCCGAAGATGGCAGTATCCTGGTACTGTTGATGTCTCTAAAAGCT
GGTGGCGTTGGGATAAATCTAACAGCTGCGTCCAATGCTTTTGTATGGATCCATGGTGG
AACCCAGCGGTAGAGGAACAAGCTGTTATGCGTATTCATCGTATAGGGCAAACCTAAGGAA
GTCAAAATCAGAAGATTCTATCGTTAAGGGAACGGTTGAAGAGAGAATGGAGGCGGTTTCAG
GCGAGGAAGCAGAGAATGATCTCTGGGGCTTTAACCGATCAAGAAGTACGAAGTGCACGT
ATAGAGGAACTCAAGATGTTATTTACCTGA

>G1526 Amino Acid Sequence (domain in AA coordinates: 493-620, 864-1006)

MGTKVSDDLSTVRSVVGSDYSDMDIIRALHMANHDPATAAINIIFDTPSFAKPDVATPTP
SGSNGGKRVDSGLKGCTFGDSSVGANHRVEEENESVNGGGEESVSGNEWFWVGCSELAG
LSTCKGRKLKSGDELVFTTFPHSKGLKPETTPGKRGFGRGRPALRGASDIVRFSTKDSGEI
GRIPNEWARCLLPVLRDKKIRIEGSCSAPEALSIMDTILLSVSVYINSSMFQKHSATSF
KTASNTAEESMFHPLPNLFRLLGLIPFKKAEFTPEDFYSKKRPLSSKDGSAIPTSLQLN
KVKNMNQDANGDENEQCISDGLDNIIVGVGDSSGLKEMETPHTLLCELRPYQKQALHWM
QLEKGNCTDEAATMLHPCWEAYCLADKRELVVYLNSTGDATIHFPSTLQMARGGILADA
MGLGKTVMTISLLLAHSWKAASTGFLCPNYEGDKVISSVDDLTSPPVKATKFLGFDKRL
LEQKSVLQNGGNLIVCPMTLLGQWKTEIEMHAKPGSLSVYVHYGQSRPKDAKLLSQSDVV
ITTYGVLTSEFSQENSADHEGIYAVRWFRIVLDEAHTIKNSKSQISLAAAALVADRRWCL
TGTPIONNLEDLYSLRLFLRIEPTWGTWAWWNKLVQKPFEEGDERGLKLVQSILKPIMLRR
TKSSTDREGRPIILVLPADARVIYCELSERDFYDALFKRSKVKFDQFVEQGVVHNYA
SILELLRLRQCCDHPFLVMSRGTAEYSDNLKLSKRFSLGKSSGLEREGKDVPEAFVQ
EVVEELRKGEQGECPICLEALEDAVLTPCAHRLCRECLLASWRNSTSGLCPVCRNTVSKQ
ELITAPTESRFQVDVEKNWVSSKITALLEELEGRLSSGSKSILFSQWTAFLDLLQIPLS
RNNFSFVRLDGLTSQQQREKVLKEFSEDGSILVLLMSLKAGGVGINLTAASNAFVMDPWW
NPAVEEQAVMRIHRIGQTKVKIRRFIVKGTVEERMEAVQARKQRMISGALTDQEVRSAR
IEBLKMLFT*

>G1543 (1..828)

ATGATAAACTACTATTTACGTACATATGCACATACACATATAAACTATATGCTCTATAT
CATATGGATTACGCATGCGTGTGTATGTATAAATATAAAGGCATCGTCACGCTTCAAGTT
TGTCCTCTTTTATATTAACTGAGAGTTTCTCTCAAACTTTACCTTTCTTCTCGATC
CTAGCTCTTAAAGAACCTAATAATTCATTGATCAAAATAATGGCGATTTTGCCGGAAC
TCTTCAAACTTGGATCTTACTATCTCCGTTCCAGGCTTCTCTTATCCCTCTCTCCGAT
GAAGGAAGTGGCGGAGGAAGAGACCAGCTAAGGCTAGACATGAATCGGTTACCGTCGTCT
GAAGACGGAGACGATGAAGAATTCAGTCACGATGATGGCTCTGCTCCTCCGCGAAAGAAA
CTCCGTCTAACCAGAGAACAGTCACGTCTTCTTGAAGATAGTTTCAGACAGAATCATACC
CTTAATCCCAAACAAAGGAAGTACTTGCCAAGCATTGTATGCTACGGCCAAGACAAATT
GAAGTTTGGTTTCAAAACCGTAGAGCAAGGAGCAAAATGAAGCAAAACCGAGATGGAATGC
GAGTATCTCAAAAGGTGGTTTGGTTTCAATTAACGGAAGAAAACACAGGCTCCATAGAGAA
GTAGAAGAGCTTAGAGCCATAAAGGTTGGCCCAACAACGGTGAACCTGCGCTCGAGCCTT
ACTATGTGTCTCGCTGCGAGCGAGTTACCCCTGCCGCGAGCCCTTCGAGGGCGGTGGTG
CCGGTCCGGCTAAGAAAACGTTTCCGCCGCAAGAGCGTGATCGTTGA

>G1543 Amino Acid Sequence (domain in AA coordinates: 135-195)

MIKLLFTYICTYTYKLYALYHMDYACVMYKYKIVTLQVCLFYIKLRVFLSNFTFSSSI
LALKPNNSLIKIMAILPENSSNLDLTISVPGFSSSPLSDESGSGGRDQLRLDMNRLPSS
EDGDDEEFSHDDGSAPPRKKLRLTREQSRLLEDSEFRQNHNLNPKQKEVLAKHLMLRPRQI
EWWFQNRARSKLKQTEMECEYLKRWFGSLTEENHRLHREVEELRAIKVGPTTVNSASSL
TMCPRCERVTPAASPSRAVVPVPAKKTFFPQERDR*

>G162 (101..619)

AGACATACAACACCAAAATCTTCTTCTTACCAACATATTCACTTTCACAGCAAAAAAAA
ACGAGAGGTTCTCTCTTATTCGTACCGTTTGTAGCAAAACAAATGGGTCGGAGAAAGATCAA
GATGGAGATGGTTTACGACATGAACACACGACAGGTTACCTTTTCAAAACGGAGGACTGG
TTTGTTCAGAAAGGCGAGCGAGTTAGCCACGCTCTGCAACGCTGAGTTGGGCATCGTTGT
CTTTTACACAGGAGGCAAGCCTTTCTCTACGGGAAACCGAATCTTGATTCTGTTGCAGA
GCGATTCTAGAGAGATATGATGATTAGACAGTGGCGATGAAGAAAAAAGTGTAATTA
CAGGCCTAAACTGAAGAGGCTGAGTGAACGTCTCGATTGTCTCAACCAAGAGGTTGAAGC
TGAGAAGGAACGAGGCGAGAAGAGTCAGGAGAAGCTTGAATCTGCTGGGGATGAGAGATT
CAAGGAGTCCATTGAGACGCTTACCTCGATGAACCTCAATGAATACAAAGATAGGCTTCA
GACAGTCCATGGTAGGATTGAAGGTCAAGTCAATCACTTGCAGGCTTCGTCTTGCCCTCAT
GCTTCTCTCCAGAAAATAGCTAGACCGACTTGTAGAGTTACATTCTATTTTTTGTATCA
GCCTACAGAACTTACCAACACATGAAAGTTATTGCTGGTGTAGAATTTTCTGTCTATAT
GGGGTGTGACTTTCTATTGACATCAAATGAAATGTACCTGGAATTTGTCTGTATTAA
TCTCAAGTGTACTTGTCTAACTTGATCAGCTTTTTCGCAAAAAAAA

>G162 Amino Acid Sequence (domain in AA coordinates: 2-57)

MGRRIKIMEMVQDMNTRQVTFSKRRITGLFKKASELATLCNAELGIVVVFSPGGKPFPSYKGP

NLDSVAERFMREYDDSDSGDEEKSGNYRPKLKRLSERLDLLNQVEAEKERGEKSQEKLE
SAGDERFKESIETLTLDLNEYKDRLQTVHGRIEGQVNHQASSCLMLLSRK*

>G1640 (168..1196)

TTCCGCCAGATCCTTCCTCTATATAAGGAAGTTCATTTTCATTTGGAGAGGTTTCGCTGACA
AGCTGCTCTAGCTTATCTGGTACCGTCGACCTCTCACTCAAGGGTCCAAAAGTGTTCCT
CTTTTTTCAGTTTCTCTTTCTCTTTTTGACAGAAGAGACCGAGAAGCAATGGGAAGGGCTC
CGTGTGTGAGAAAATCGGGTTGAAGAGAGGGAGATGGACAGCCGAGGAAGATGAGATCC
TCACCAAGTATATTACAGACCAATGGTGAAGGTTCTTGGCGATCTTTGCCTAAGAAAGCTG
GATTGTTGAGATGTGGAAAGAGCTGTAGACTAAGGTGGATAAACTACTTAAGAAGAGACT
TAAAAAGAGGAAATATTACTTCCGACGAAGAAGAAATAATCGTCAAGTTGCATTCCCTTC
TCGGCAACAGATGGTCACTTATTGCAACACATCTACCAGGAAGAACAGACAACGAAATTA
AAAACATATTGGAACCTCACATCTCAGCCGCAAAATCTATGCCTTCACTGCCGTTTCCGGAG
ATGGACACAATCTACTCGTCAACGATGTAGTCTTGAAGAAATCTTGTTTCATCGTCTTCTG
GAGCCAAGAACAATAACAAGACCAAGAAGAAGAAGGAAGGACTAGTAGGTCATCCA
TGAAGAAACACAAGCAAAATGGTGACGGCTCACAATGTTTCTCACAACCTAAGGAGCTAG
AGAGTGATTTCACTGAGGGAGGGCAAAATGGTAATTTTGAAGGAGAGTCTTTGGGGCCTT
ATGAGTGGTTGGATGGTGAGTTAGAACGGCTCTTGAGTAGTTGTGTCTGGGAATGCACTA
GTGAAGAGGCTGTGATTGGAGTAAATGATGAAAAGGTGTGTGAGAGTGGGGACAATAGTA
GTTGTTGTGTTTAATTTGTTTGAAGAAGAAACAAGGAAGCGAGACAAGATTGGTCACGTAG
GAATCACAGAGGTTGATCATGATATGACGGTGGAAGAGAAAGAGAGGGGAAGTTTTTTAA
GTTCAATTCAAATGAAAATAATGATAAAGATTGGTGGGTTGGTCTATGTAATTCCTCAG
AAGTTGGGTTTGGGGTTGATGAGGAGTTGCTTGATTGGGAGTTTCAAGGTAATGTCACTT
GTCAAAGTGATGATCTATGGGATCTCTCAGATATTGGAGAGATAACATTGGAGTGATTGT
ACCGAGCAAGTGGATTGGCGGCCGCTCTAGACAGGCCTCGTACCGATCTCTAGCTAGAG
CTTTCGTTTCGTATCATCGGTTTCGACAACGTTTCGTCAAGT

>G1640 Amino Acid Sequence (domain in AA coordinates: 14-115)

MGRAPCCEKIGLKRGRWTAEEDEILTKYIQTNGEGSWRSLPKKAGLLRCGKSCRLRWINY
LRRDLKRGNIISDEEEIIVKLHSLGNNRWSLIATHLPGRTPDNEIKNYWNSHLRKYAFT
AVSGDGHNLVNDVVLKKSCSSSSGAKNNNKTKKKKKGRTSRSSMKHKQMVTASQCFSQ
PKELESDFSEGGQNGNFEGESLGPYEWLDGELERLLSSCVWECTSEEAVIGVNDKVCES
GDNSSCCVNLFFEEQSGSETKIGHVGITEVDHDMTVEREREGRSFLSSNSNENNDKDWVGL
CNSSEVGFVDEELLDFEQNVTCQSDDLWDLSDIGEITL*

>G1644 (1..348)

ATGAAATTGATTGATTGGAAAGACTGTGCTTTGATGACTTACACCGAAGTCAATTTGGGT
TTCTGCAATGTTTTAATGTTGATCTGCAGGAGGACTAGTGGACCTATGAGACGAGCAAAA
GGTGGTTGGACTCCAGAGGAGGATGAGACACTTAGACGAGCAGTTGAAAAGTATAAGGGG
AAGAGGTGGAAGAAAATAGCGGAATTTTCCAGAGAGAACACAAGTCCAATGCTTGCAC
AGGTGGCAGAAAGTTCTTAATCCAGAGCTTGTAAAGGACCTTGGACTCAAGAGGTTCTC
TTATCATTTTCATGTTCTGAACTTTTTTTGGTTTTTCATTTTACGTAA

>G1644 Amino Acid Sequence (conserved domain in AA coordinates: 39-102)

MKLIDWKDCALMTYTELILGFCNVLMILICRRTSGPMRRAKGGWTPEEDELRRAVEKYKG
KRWKKIAEFFPERTQVQCLHRWQKVLNPELVKGPWTQEVLLSFSCSETFFGFHFT*

>G1646 (34..786)

GATCTTTTGATCCAATCACAAGGCAAAGATCCAATGGACAATAACAACAACAACAAC
CAGCAACCACCACCAACCTCCGTCTATCCACCTGGCTCCGCCGTCACAACCGTAATCCCT
CTCCACCATCTGGATCTGCATCAATAGTCACCGGAGGAGGACGACATACCACCACCTC
CTCCAGCAACAACAGCAACAGCTTCAAATGTTCTGGACATACCAGAGACAAGAGATCGAA
CAGGTAAACGATTTCAAAAACCATCAGCTCCCTCTAGCTCGTATCAAAAAAATCATGAAA
GCTGATGAAGATGTGCGTATGATCTCCGCCGAAGCACCGATTCTTTCGCGAAAGCTTGT
GAGCTTTTCATTCTCGAATTACGATTAGATCTTGGCTTCACGCTGAAGAGAACAAACGT
CGTACGCTTCAGAAAAACGATATCGCTGCTGCGATTACTAGAACCGATATCTTCGATTTC
CTTGTTGATATTGTTCTAGGGAAGAGATCAAGGAAGAGGAAGATGCAGCATCGGCTCTT
GGTGGAGGAGGTATGGTTGCTCCCGCCGAGCGGTGTTCTTATTATTATCCACCGATG
GGACAACCGCGGTTCTCGGAGGGATGATGATTGGAAGACCGGCGATGGATCCTAGCGGT
GTTTATGCTCAGCTCCTTCTCAGGCATGGCAAAGCGTTTGGCAGAATTCAGCTGGTGGT
GGTGATGATGTCTTATGGAAGTGGAGGAAGTAGCGGCCATGGTAATCTCGATAGCCAA
GGGTAAGTGAATTCTAGTAG

>G1646 Amino Acid Sequence (domain in AA coordinates: 72-162)

MDNNNNNNNQPPPTSVYPPGSAVTTVIPPPSGSASIVTGGGATYHHLLQQQQQLQMF
WTYQROEIEQVNDFFKNHQLPLARIKKIMKADEDVRMISAEAPILFAKACELFILELTIRS
WLHAEENKRRTLQKNDIAAAITRTDIFDFLVDIVPREEIKEEBDAASALGGGGMVAPAAS
GVPIYYPPMGQPAVPGGMMIGRPAMDPSGVYAQPPSQAWQSVWQNSAGGGDDVSYGSGGS
SGHGNLDSQG*

>G1672 (239..1399)

CCATTCTGACGTCCGGGATGACGCACAATCCCCTATCCTTCGCAAGACCCTTCCTCTA
TATAAGGAAGTTCATTTCATTTGGAGAGGACACGCTGACAAGCTGACTCTAGCAGATCTG
GTACCGATCACTCCCGTCTTTATCAAATTCCTTCTCCTCTTACATTTTCCCTATCCAATC
GATCTCACGCAGATCTGATCAATTTCTCATCAAATCATTTAGAGATCAAAAGAAAATAT
GAAGAATAGTAAATGTAACCTCATAGATTCAAAGCTCGAAGAACATCATCATCTTTGCGG
ATCAAAACATTGTCTGGATGTGGTTCGCATGATTCAAGCTGCTACTAAACCAAATTGGGT
TGGATTGCGCGGAGGAGTGAAAATTCGATCCGACAGATCAAGAACTTATAGAACATTTAGA
AGCAAAAGTGAAGGAAAAGAAGAAAATAAGAAATGGTTCGTCTCATCCACTTATAGA
TGAATTTATTCCCACCATGATGGAGAAGATGGAATATGTTACTCATCCTCAGAAGCT
TCCAGGGGTGACAAGAGATGGCTTGAGCAAACACTTCTTCCACAAACCATCAAGAGCTTA
CACAAACCGGAACAAGAAAACGACGTAAAATAAATCAAACCGATCAGACTCTGAGTTAAC
CGGATCATCAGAAACAGGTGGCACAACCGGCAAAACAAGACCGGTTATGATCAACGG
TCAACAAAGAGGATGCAAGAAGATATTAGTACTCTACACAACTTCGGCAAGAATCGTCG
ACCGGAGAAAACAAAATTTGGGTGATGCATCAATATCATTTAGGGATTATGAGGAAGAGAG
AGAAGGAGAATTTGTGGTCTCCAAGATATTTTATCAGACACAACCAAGACAGTGTGTTAG
TAATACTAATTGGTCTGATCACCATGGTTCCAAGGACGTGATCGGAATTGGTGTGCGAGA
TGAGATTTCCAGCGTAGCTGCCACGTTGCAGAGTCTTGGCTCCGGTGACGTCGTTTCTAG
GGTTAATATGCATCCCATACAAGATCCTTTGATGAGGGGACAGCCGAAGCTTCAAAGGG
AAGAGAGAACCAGCATGTGTCTGGCAGGTGCGAGGAAGTACATGATGGGATCATAACATC
ATCAATGTCTATCTCATATGATTCATGATCATCATAATCAACATCATCAAATCGGAGA
TAGAAGAGAATTTACATGTCATCATCATATCCCCTGACCCCTACTATCACATCACAAACA
TGAGTCAATCTTCCATGTTACAAGTACTATGCCCTTTTCAGCGGCAGCAATTAAGGGGTCG
GTCGTCTGGTTCGGGATTAGAAGACCTAATTTATGGGTTGTACCACAGCTACGTGTACAGA
AGACAATAATCACAAATGATTAAATTCGCAGGAGCATTTCAGAAGCAAACCTCAGCGAAA
TGCAGAGTGGTTAACGTTTCCACAATTCTGGAACCAAGCCGAATCAGATGATCAAACCG
AAGATTTTAACAGAACCAAAAGGAAGCAGAGAAAATCTTGCAAAAAGCTCCTGCTTAGCTG
TTGATCAATGCCGGAATGCTGAGCTATGACTGACTAGTCTCTGCCATTTAACTTACAAT
ATCACCAGAGGTTGCGATGAATGTTGATTCGCTCAAAGGAGAGCGGCCGCTCTAGACAGG
CCTCGTACCG

>G1672 Amino Acid Sequence (conserved domain in AA coordinates: 41-194)

MKNSKCNLIDSKLEHHHLCGSKHCPGCGRMIQAATKPNWVGLPAGVKFDPDQELIEHL
EAKVKGKEENKKWSSSHPLIDEFIPTIDGEDGICYTHPQKLPGVTRDGLSKHFFHKPSRA
YTTGTRKRRKIIQTDHDSLTGSSETRWHKTGKTRPVMINGQQRGCKKILVLYTNFQKNR
RPEKTNWVMHQYHLGINEEEREGLVVSIFYQTQPRQCVSNTNWSDDHHSKDVIGIGVG
DEISSVAATLQSLGSGDVSVRVMHPHTRSFDEGTAEASKGRENQHVSGTCEEVHDGIIT
SSMSSHHMIHDHNQHQQIGDRREFHMSSTYPMTPTITSQHESIFHVTSTMPFQRQQLRG
RSSGSGLEDLIMGCTTATCTEDNNHK*

>G1677 (24..1037)

CAGTACTAATTCGTGTGTGTTAATGGTTCCTAGTTATGGATGATGAAGAGAGTAACAACG
TTGAAAGATATGACGACGTCGTATTGCCAGGGTTTAGGTTCCATCCCCTGATGAAGAAC
TCGTAAGTTTCTACTTTGAAACGGAAGGTTTACACAAATCTCTTCCCTTTGATCTCATCA
AGAAAGTCGACATTTACAAATACGATCCATGGGACCTCCAAAGCTTGACGCGATGGGGG
AAAAAGAGTGGTACTTTTATTGTCTTAGAGACAGGAAATACCGCAACAGCACAAAGACCTA
ACCGAGTAACCTGGAGGTGGCTTCTGGAAGCAACCGGAACAGACCGGCCTATATACTCAT
TGGACTCCACTCGATGCATCGGTTTGAAGAAATCACTTGTGTTCTACCGTGGTCGAGCTG
CTAAAGGAGTCAAAACCGATTGGATGATGCATGAATTTCTGTCCTCCTCTCTCTGACT
CTCATCACTCATCATATCCCAATTACAATAACAAGAAGCAACACCTTAACAATAACAACA
ACAGCAAGGAGCTTCCTTCAAACGATGCTTGGGCGATATGTAGAATATTTAAGAAGACAA
ATGCAGTATCCTCACAAAGATCAATCCCACAATCTTGGGTTTATCCAACGATTCTTGACA
ACAATCAACAGTCACACAACAACACCGCAACTCTCTTAGCTTCATCAGACGTTCTCAGCC

ACATATCAACAAGACAAAACCTTTATTCTCTTCCAGTCAACGAACCCGCAAGCTTCACAG
AATCAGCTGCTTCTTACTTCGCGTCTCAGATGCTCGGAGTCACGTACAATACAGCCAGAA
ACAACGGAACAGGGGATGCTCTGTTTCTGAGAAACAATGGAACAGGGGATGCTCTGGTTC
TGAGCAACAATGAGAATACTACTTCAACAACCTTGACTGGAGGGTTGACTCATGAGGTTT
CGAATGTAAGATCAATGGTGATGGAGGAGACTACGGGGAGTGAGATGTCGGCGACGTCGT
ATTCCACTAACAATTAAGATCATAGTACTATTAACACTGAATTAGTGTAGACGTTGATC
ATCGCTAATATGTATTAATTTTTCTTGTCTTACTATAAACGAAAAA

>G1677 Amino Acid Sequence (conserved domain in AA coordinates: 17-181)

MVLVMDDEESNNVERYDDVLPGRFRHPTDEELVSFYLRKVLHKSPLFDLIKKVDIYKY
DPWDLPLKLAAMGEKEWYFYCPDRKRYRNSRPNRVTTGGFWKATGTDRIYSLDSTRCIG
LKKSLVFIYRGRAAKGVKTDWMMHEFRLPSLSDSHSSYPNYYNNKKQHLNNNNNSKELPSN
DAWAICRIFKKTNVSSQORSIPQSWVYPTIPDNNQQSHNNTATLLASSDVLISHISTRQNF
IPSPVNEPASFTESAASYFASQMLGVTYNTARNNGTGDALFLRNNGTGDALVLSNNENNY
FNNLTGGLTHEVPNVRSMVMEETTSEMSESTSYSTNN*

>G1765 (139..966)

TCCTTCGCAAGACCTTCTCTTATATAAGGAAGTTCATTTTCAATTTGGAGAGGACACGCTG
ACAAGCTGACTCTAGCAGATCTGGTACCGTCGACAAGAATGACTTGATTGGTGTCTAAA
GAGATCGATGTAGTGAAGATGAGTGGCGAAGGTAAGTGAAGGATCATGAAGAAGAA
AACGAAGCACCACCTTCTGGGTTTCAGGTTTCATCCGACGGATGAAGAGCTTTTAGGATAC
TATCTTCGAAGAAAAGTAGAGAACAAAACCATCAAACCTCGAAGTATCAAACAGATCGAT
ATCTATAAGTACGATCCTTGGGATCTTCCAAGAGTGAGCAGCGTCGGAGAAAAGGAGTGG
TACTTCTTCTGCATGAGAGGTAGGAAATACAGGAATAGCGTTCGACCAAACCGAGTGACC
GGTTCAGGTTTCTGGAAGCCACTGGTATTGATAAACCGGTTTACTCCAATCTTGACTGT
GTTGGTCTCAAGAAATCTCTGGTTTACTATCTTGGTTCAGCCGGTAAAGGCACCAAAC
GATTGGATGATGCATGAATTCGCCCTCCCTCCACCACGAAAACCGACTCTCCAGTCAA
CAAGCAGAGGTATGGACACTTTGCAGAATCTTCAAACGAGTCACATCTCAAAGAAACCCA
ACCATCTTACCACCAAACCGAAAACCGGTTATCACTTTAACCGACACTTGTCTAAGACC
AGCAGCTTAGATTCCGACCACACGAGCCACCGTACAGTAGATTCCATGTCCCACGAGCCG
CCGCTTCCACAGCCACAGAATCCTTATTTGGAACCAACATATAGTTGGTTTTAATCAACCG
ACATATACTGGTAATGATAATAACCTCCTGATGAGTTTCTGGAACGGCAACGGTGAGAT
TTCATAGGAGACTCAGCAAGTTGGGATGAACCTTAGATCTGTTATAGATGGCAACACTAAA
CCCTAGTAATAAAGTTTCTTTTTCAGCTTTGTACAAAAGATAAAACAAACGGCAACC
GCTCTAGACAGGCCCTCGTACCGGGATCCTCTAGCTAGAGCTTTCGTTTCGTATCATCGGT
TTCGACAACGTTCTGT

>G1765 Amino Acid Sequence (conserved domain in AA coordinates: 20-140)

MSGEGNLGKDHEEENEAPLPGRFRHPTDEELLGYLRRKVENKTIKLELIKQIDIIKYDP
WDLPRVSSVGEKEWYFFCMRGRKYRNSVRPNRVTTGSGFWKATGIDKPVYSNLDVCVLKKS
LVYYLGSAGKGKTDWMMHEFRLPSTTKTDSAPAQAEVWTLCRIFKRVTSQRNPTILPPN
RKPVITLTDTCSTSSLDSDHTSHRTVDSMSHEPPLPQPQNPYWNQHVGFNPQPTYTEND
NNLLMSFWNGNGDFIGDSASWDELRSVIDGNTKP*

>G1777 (97..1878)

CTCGTACTTTATCACCTCCGTCGTTCTATAATACTCTCTTCCGTCAATCATATCATTGT
CGACAATTTTATTCTGATCAGTTTAAAAATTGATCCATGGATGATAATTTAAGCGGCGAG
GAAGAAGATTACTATTACTCTCCGATCAGGAATCTCTCAACGGGATTGATAATGATGAA
TCCGTTTCGATACCTGTTTCTTCCCGATCAAATACTGTCAAGGTTATTACGAAGGAATCA
CTTTTGGCTGCACAGAGGGAGGATTGCGGAGAGTGATGGAATTGTTATCGGTTAAGGAG
CACCATGCTCGGACTCTTCTTATACATTACCGATGGGATGTGGAGAAGTTGTTTGCTGTT
CTTGTTGAGAAAGGGAAGATAGCTTGTCTTCTGGTGTGGTGTACACTTCTTGAAAAC
CAAAGTTGTGATTCTTCCGTTTCTGTTTCTTCTCGATGATGAGTTGTGATATCTGCGTA
GAGGATGTACCGGTTATCAGCTGACAAGGATGGACTGTGGCCATAGCTTTTGCAATAAC
TGTGAGACTGGGCATTTTACTGTAAAGATAAATGAAGGTCAGAGCAAAAGGATTATATGC
ATGGCTCATAAAGTGAATGCTATTTGTGATGAAGATGTTGTGAGGCTCTAGTTAGTAAA
AGCCAACAGATTTAGCTGAGAAGTTTGATCGTTTCTTCTTGAGTCGTATATCGAAGAT
AACAAAATGTTGGAAGTGGTGTCCGAGTACTCTCATTTGTTGGAATGCCATACGTGTTGAG
GATGACGAGCTCTGTGAGGTTGAATGCTCTTGTGGTTTGAGTTCTGTTTCAGTTGTTCA
TCTCAAGCTCACTCCCTTGTCTTGTGTGATGTGGGAATATGGAGAAAGAAGTGCCTT
GATGAGTCCGAGACTGTTAATTGGATAACTGTTTACACAAAGCCGTGTCCCAAATGTCAC

AAGCCTGTTGAAAAGAATGGTGGATGCAATCTCGTGACTTGTCTTTGTGACAATCTTTT
TGTTGGTTGTGTGGTGAAGCTACTGGAAGGGACCACACTTGGGCTAGAATCTCGGGTCAT
AGTTGTGGTCGGTTCCAAGAAGATAAAGAGAAACAAATGGAGAGAGCGAAAAGGGATCTC
AAGCGGTATATGCATTATCATAACCGATACAAAGCACATATCGACTCTCCAAGCTAGAG
GCTAAGCTTAGTAATAATATTAGTAAAAAGGTGTCTATTTCAGAAAAGAGGGAGTTACAA
CTTAAAGACTTCAGCTGGGCTACCAATGGACTCCATCGGTTATTTAGATCAAGACGAGTT
CTTTCATATTACATACCTTTTCGCATTTTACATGTTTGGAGATGAGCTGTTTAAAGATGAG
ATGAGCTCTGAGAAAGAGAAATAAAACAAAATCTGTTTGGAGATCAGCAGCAGCAGCTT
GAGGCTAATGTTGAGAACTTTCTAAGTTCTTGGAGGAACCTTTTGATCAATTTGCTGAT
GATAAGGTCATGCAGATAAGGATTCAAGTCATCAATTTGTGAGTTGCGGTCGATACACTC
TGCGAAAATATGTATGAATGCATTGAGAATGACTTGTGGGTTCTCTGCAACTTGGCATC
CACAACATTACTCCATACAGATCAAAACGGCATAGAACGAGCATCTGATTTTTATAGTTCC
CAGAATTTCAAGGAAGCTGTTGGTCAGTCTTCCGATTGTGGATGGACGTCAGGCTCGAT
CAAGCTTTGGAGTCAGGGAAGTCGGAAGACACAAGTTGCTCTTCCGGGAAGCGTGCTAGA
ATAGACGAAAGTTACAGAAACAGCCAAACACCTTACTAGATTTAACTTGCCAGCGGAA
GCCATTGAGCGGAAATGAACACTTATCCTTCTTCACTCCCAATAACACCTTTTGTGCC
AAATAAAGTGTGTTACCCGGATATTTATAGCTCTAAACCCAATCCCCTCTGCTTAATTTG
TCAGTGACCTTACCTAACCTCTTCA

>G1777 Amino Acid Sequence (domain in AA coordinates:124-247)
MDDNLSGEEEDYYSDDQESLNGIDNDESVSIPVSSRSNTVKVITKESLLAAQREDLRRV
MELLSVKEHHARTLLIHYRWDVEKLFPAVLVEKGKDSLFSGAGVTLLNQSCDSSVSGSSS
MMSCDICVEDVPGYQLTRMDCGHSFCNNCWGTGHFTVKINEGQSKRIICMAHKCNAICDED
VVRLVSKSQPDLAELKDFRFLLESYIEDNKMVKWCPSTPHCGNAIRVEDDELCEVECSG
LQFCFSCSSQAHSPCSCVMWELWRKKCFDESETVNWITVHTKPCPKCHKPVEKNGGCNLV
TCLCRQSFCLWLCGEATGRDHTWARISGHSCGRFQEDKEKQMERAKRDLKRYMHYHNRYKA
HIDSSKLEAKLSNNISKVVISSEKRELQKDFSWATNGLHRLFRSRRVLSYSYPFAFYMF
GDELFPKDEMSSEEREIKQNLFEQQQLEANVEKLSKFLEPPDQFADDKVMQIRIQVIN
LSVAVDTLCENMYECIENDLLGSLQLGIHNITPYRSNGIERASDFYSSQNSKEAVGQSSD
CGWTSRLDQALESKSEDTSCSSGKRARIDESYRNSQTLLDLNLPAAIERK*
>G1793 (59..1783)

AGTGATTATTGATTAACCCAAACACAAAATAAACAGATTTGACTCAAAAAGAAGAAAAT
GAATTCTAACAACCTGGCTTGGCTTTCCTCTTTACCGAACAACCTTCTTTGCTCTCTCA
TGAATACAACCTTGGCTTGGTCAGCGACCATATGGACAACCTTTTCAAACACAAGAGTG
GAATATGATCAATCCACACGGTGGAGGAGGAGATGAAGGAGGAGAGGTTCCAAAAGTGGC
CGATTTTCTCGGTGTGAGCAAACCGGACGAAAACCAATCCAACACCTAGTAGCTTACAA
CGACTCAGACTACTACTTCCATACCAATAGCTTGATGCCTAGCGTCCAATCAAACGATGT
CGTTGTAGCAGCTTGTGACTCCAATACTCCTAACAACAGTAGCTATCATGAGCTTCAAGA
GAGTGCTCACAATCTACAGTCACTTACTTTGTCCATGGGGACCAACGCTGGTAATAATGT
TGTAGACAAAGCTTACCATCCGAGACCACCGGGGATAACGCTAGCGGTGGAGCACTAGC
CGTTGTTGAGACGGCCACGCCAAGACGTGCATTGGACACTTTCGGACAACGAACCTCGAT
CTATCGTGGTGTCAAGACATCGATGGACTGGTGCATATGAGGCTCATCTATGGGATAA
TAGTTGTAGAAGGGAAGGCCAGTCTAGGAAAGGAAGACAAGTTTACTTGGGTGGATATGA
CAAAGAAGATAAAGCAGCAAGATCATATGATCTAGCTGCCTTAAGTACTGGGGTCCTTC
AACTACTACTAATTTCCCATTAACAACCTACGAGAAAGAGTAGAGGAAATGAAGCACAT
GACGAGACAAGAGTTCTGTGGCTGCCATTAGAAGGAAAAGTAGTGATTTTCGAGAGGCGC
TTCGATGTATCGAGGAGTTACAAGGCATCAACAACATGGAAGATGGCAAGGATCGG
CCGAGTCGCCGGAACAAGACCTTACTTGGGAACCTTTAGCACTGAGGAAGAAGCAGC
AGAAGCTTACGATATAGCTGCAATAAAGTTTAGAGGACTTAATGCAGTGACCAACTTCGA
GATCAACCGGTACGACGTGAAAGCCATTCTAGAGAGTAGCACTCTTCCATCGGAGGAGG
CGCAGCTAAACGGCTCAAAGAAGCTCAAGCTCTTGAGTCTTCAAGGAAACGCGAGGCGGA
GATGATAGCCCTTGGTTCAAGTTTCCAGTACGGTGGTGGCTCGAGCACAGGCTCTGGCTC
CACCTCATCAAGACTTCACTTCAACCTTACCCTCTAAGCATTCAACAACCATTAGAGCC
TTTTCTATCTCTTCAGAACAAATGACATCTCTCATTACAACAACAATGCTCAGGATTC
CTCCTCTTTAATCACCATAGCTATATCCAGACACAACCTTCATCTCCACCAACAGACCAA
CAATTACTTCAGCAACAGTCGAGCCAGAACTCTCAGCAGCTCTACAATGCGTATCTTCA
TAGCAATCCGGCTCTGCTTCATGGACTTGTCTCTACCTCTATCGTTGACAACAATAATAA
CAATGGAGGCTCTAGTGGGAGCTACAACACTGCAGCATTTCTTGGGAACCACGGTATTGG

TATTGGGTCCAGCTCGACTGTTGGATCGACCGAGGAGTTTCCAACCGTTAAAAACAGATTA
CGATATGCCCTTCCAGTGATGGAACCGGAGGGTATAGTGGTTGGACCAGTGAGTCTGTTCA
GGGGTCAAACCTGGTGGTGTCTTTCACTATGTGGAATGAGTAAACAAGGATCTCTTTCTT
GCGGCACAAGGAATGGGT

>G1793 Amino Acid Sequence (conserved domain in AA coordinates: 179-255, 281-349)

MNSNNWLGFPLSPNNSSLPPHEYNLGLVSDHMDNPFQTQEWNMINPHGGGGDEGGEVPKV
ADFLGVSKPDENQSNHLVAYNDSYYFHTNSLMPVQSNVVAACDSNTPNNSSYHELQ
ESAHLNQLSLTSMGTTAGNNVVDKASPSETTGDNASGGALAVVETATPRRALDTFGQRTS
IYRGVTRHRWTGRYEAHLWDNSCRREGQSRKGRQVYLGGYDKEDKAARSYDLAALKYWGP
STTTNFPITNYEKEVEEMKHMTRQEFVAAIRRKSSGFSRGASMYRGVTRHHQHGRWQARI
GRVAGNKDLYLGTFTSEEEAAEYDIAAIKFRGLNAVTFEINRYDVKAILESSTLPIGG
GAAKRLKEAQALESRRKREAEMIALGSSSQYGGSSSTGSGSTSSRLQLQPYPLSIQOPL
PFLSLQNNNDISHYNNNNAHDSSSFNHSYIQTQLHLHQQTNNYLQQQSSQNSQQLYNAYL
HSNPALLHLGVSTSIVDNNNNNGSSGSYNTAFLGNHIGIGSSSTVGSTEEFPTVKTD
YDMPSSDGTGGYSGWTSESVQGSNPGGVFTMWNE*

>G180 (54..629)

GTAATTACGATCTACAACAAGTGACATCGTCGTCGACGACGATTCAAGAGAATATGAAC
TCCTCGTTCCTTTGAAGAAACCAATGTCTTAACCTTTTCTCTTCTTCTTCTCTCTT
CTCTTTCTTCTCTCTTCTTCCCATTCACAACCTCTCTCTCCACTACTACTACTCATGCAC
CTCTAGGGTTTTCTAATAATCTTCAGGGTGGAGGACCCCTGGGATCAAAGGTGGTTAATG
ATGATCAGGAGAATTTGGAGGTGGAACATAACAATGATGCTCATTCTAATTCTTGGTGA
GATCAAATAGTGAAGTGGAGATATGAAGAACAAAGTGAAGATAAGGAGGAAACTAAGAG
AGCCAAGATTCTGTTTCCAAACCAAAGCGATGTTGATGTTCTTGACGATGGCTACAAAT
GGCGTAAATATGGTCAGAAAGTCGTCAAGAACAGCCTTCACCCAGGAGTTATTACAGAT
GCACACACAACAACGTAGGGTGAAAAAGAGAGTGAGCGACTATCGGAAGATTGTAGAA
TGGTGATTACTACTTACGAAGGTCTGCACAAACCATTCCTCTCTGATGACTCCACTTCTC
CTGACCATGATTGTCTCTCTTCTTTTAAACATCTCTTTCTATATATCTATATATAGACAG
TTATATGTGCACATATAGATGTGTGATATATTGCATATTTGATATTGCATGTGTTTTCA
AGAGTATGTCATCAGATGTTATGCATATATTCTTGACTTGTGCTTATAGTATACATATG
TAATAATATATATTGACATTGGTAGTTCATTTCTGTTCAAACAAAAAAAAAAAAAA

>G180 Amino Acid Sequence (domain in AA coordinates: 118-174)

MNFLVPFEETNVLTFSSSSSSSLSSPSFPIHNSSSTTTTHAPLGFNNLQGGGPLGSKV
VNDDQENFGGTTNNDASHNSWWSNSGSGDMKNKVKIRKRLREPRFCFQTKSDVDVLDG
YKWRKYGQKVVKNSLHPRSYYRCHTHNCRVKRVERLSEDCRMVITTYEGRHNHIPSDDS
TSPDHDCLSSF*

>G192 (63..959)

CTTTTTTCTTCTCTCTCCTCAGAGATTGGAAGCTTTTTGTCTCCCTGAGTAACCAAATT
CAATGGCCGACGATTGGGATCTCCACGCCGTAGTCAGAGGCTGCTCAGCCGTAAGCTCAT
CAGCTACTACCACCGTATATCCCCGGCGTTTCATCTCACACAAACCTATATTCACCG
TCGGACGACAAAGTAATGCCGTCTCCTTCGGAGAGATTGAGATCTCTACACACCGTTCA
CACAAGAATCTGTCGTCTCTTCGTTTTCTGTATAAACTACCCAGAAGAACCTAGAAAGC
CACAGAACCAGAAACGTCCTCTTTCTCTCTCTGCTTCTCCGGTAGCGTCACTAGCAAAC
CCAGTGGCTCCAATACCTCTAGATCTAAAAGAAGAAAGATACAGCATAGAAAGTGTGCC
ATGTAGCAGCAGAAGCTTTAAACTCCGATGCTCTGGGCATGGCGAAAGTACGGACAGAAAC
CCATCAAAGGTTACCATATCCAAGAGGATACTACAGATGTAGTACATCAAAGGTTGTT
TAGCCCGTAAACAAGTGGAGCGAAATAGATCCGACCCGAAGATGTTTATCGTCACTTACA
CGCGGAGCATATATCATCCAGCTCCGACACACCGTAATTCTCTCGCCGGAAGCACACGTC
AGAAACCATCCGATCAACAGACGAGTAAATCTCCGACGACCACTATTGCTACTTATTCAT
CGTCTCCGGTGACTTCAGCCGACGAATTTGTTTTGCCTGTTGAGGATCATCTAGCGGTGG
GAGATCTTGACGGAGAAGAAGATCTGTTATCTTTGTCGGATACGGTGGTTAGCGATGATT
TCTTCGATGGGTAGAGGAATTTCGACGCCGAGATAGCTTTTCCGGGAACCTCGGCTCCGG
CGAGTTTTGATCTCTCTTGGGTTGTGAACAGTGCCGCCACTACCACCGAGGAATATGAT
TAGATTACGACGGCTTAGAATACTCTTATTAGGACAGATTATAGGATTAAGGAATTATT
CTCGGAGCATATGTAAAAATAGGATAAAAAGAAAATGTTCTTTGTTACTTTTTTTTCGGGT
TTCTTCTTATTGTTTCTAAACATCTTAGAAAAAATTTAATTGTATATTCCTTAAGCTCGA
TACATCTTGTTTTAAAAAATAAAAAAAAAA

>G192 Amino Acid Sequence (domain in AA coordinates: 128-185)

MADDWDLHAVVRGCSAVSSSATTTVYSPGVSSHTNPIFTVGRQSNVSVFGEIRDLYTPFT
QESVVSFSCINYPPEPRKPQNQKRPLSLSASSGSVTSKPSGNTSRSKRRKIQHKKVCH
VAAEALNSDVWAWRYGQKPIKGSYPYRGYYRCSTSKGCLARKQVERNRSDPKMFIVTYT
AEHNHPAPTHRNSLAGSTRQKPSDQQTSSPTTTIATYSSSPVTSADFLVPVEDHLAVG
DLDGEEEDLLSLSDTVVSDDFDGLLEFAAGDSFSGNSAPASFDLSWVNSAATTTGGI*
>G1948 (18..1118)
AAAAGGCTTCTTGCCCATGGATACTTGTGCTCTAGTAATCCATCAGTCTCTGTCTCGCA
TCAAACCTTCTCTCCCAAATCTTCTTCTTCTTCTTCTGCTTTCTCCCTGAATCCT
TACCGATCAGACGGATCGAGCTGTGTTTCCGAGGAGCTATATGTGCCGCCGTACAAAGAA
ACTACGAAGAAACGACCTCCTCCGTGGAAGAGGCAGAGGAAGATGATGAGTCATCATCAT
CGTACGGAGAAGTGAACAAGATCATTTGGAAGCCGAACGGCGGGGAAGGAGCCATGGAGT
ACCTTATCGAGTGAAGGACGGCCATTCTCCGTCTGCGGTTCATCGAGCTACATCGCAG
CAGACGTAGTGTGCGAGTACGAGACACCTGGTGGACGGCAGCTAGAAAAGCCGACGAGC
AGGCCCTGTACAGCTCCTGGAGGACCGAGACGTGATGCCGTGGACGAAAACGGCCGGA
CGGCTCTGCTTTTCTGCGGAGGTCTGGGGTGGGACAAAGTGCCTAAGGCTTCTGGCGGAGG
CTGGAGCCGATCTCGACCACCGAGACATGAGGGGAGGCTTGACGGCGCTGCACATGGCGG
CTGGTTACGTGAGGCCGGAGGTGGTGGAGGCGCTGGTGGAGCTGGGAGCTGATATTGAAG
TGGAAGACGAGAGGGGTTAACGGCGTTGGAACAGCGAGGGAGATTCTGAAGACGACGC
CGAAGGGGAATCCGATGCAGTTCCGGAGGAGAATTGGGTTAGAGAAAGTATCAATGTCC
TGGAAGGACAAGTGTTCGAGTACGCCGAGGTGGATGAGATCGTAGAGAAACGAGGGAAAG
GCAAAGACGTTGAATATCTGGTCAGATGGAAGACGGTGGAGATTGCGAGTGGGTGAAAG
GTGTACACGTGGCGGAAGATGTGGCTAAGGACTACGAGGATGGGCTGGAGTACGCTGTAG
CGGAGAGTGTGATCGGGAAGAGGTGGGAGACGATGGGAAGACCATCGAGTATCTTGTCA
AATGGACTGATATGTCTGATGCCACTTGGGAGCCTCAGGACAATGTCGACTCTACTCTTG
TTCTACTCTACCAACAACAACCAATGAATGAATGATTGATTTTGTATGATTACATTCT
TCTCAATTTGCTTCTTTCTCATATGTGTTGGTTCATCTGACCGGTTCCGTTGGTACGTAC
CGGTACATTTTTCATTTTCTTTTAAGATGTGATCTTGTATGGTTTTTGGCCTTTTGGGGACA
CTATTGATTTTATATCCATGCTTTGAATTTTGCTTCCCTTTTGGGGAGATTTCATGAAA
>G1948 Amino Acid Sequence (domain in AA coordinates: entire protein)
MDTCALVIHQSLSRILKSPKSSSSSSSAFSPESLPIRRIELCFRGAICAAVQRNYEET
SSVEAEEDDESSSSSYGEVNIIGSRTAGEGAMEYLIENKDGHSWVPSYIAADVSE
YETPWWTAAARKADEQALSQLEDRDVEDENGRTALLFVAGLGSDKCVRLLEAGADLD
HRDMRGGTLALHMAAGYVRPEVVEALVELGADIEVEDERGLTALELAREILKTPKGNFM
QFGRRIGLEKVINVLEGOVFEYAEVDEIVEKRGKGDVEYLVVRWKDGGDCEWVKGVHVAE
DVAKDYEGLLEYAVAESVIGKRVGDDGKTI EYLVKWTDMSDATWEPQDNVDSTLVLLYQQ
QQPMNE*
>G2123 (1..657)
ATGAGAAAAGTATGTGAGCTTGATATAGAGCTAAGTGAAGAGGAAAGAGACCTACTAACA
ACTGGATACAGAATGTCTATGGAGGCTAAGAGAGTTTCATTGAGAGTAATATCATCCATT
GAAAAAATGGAAGACTCGAAAGGAAACGACCAAAATGTGAAACTGATAAAGGACAACAA
GAAATGGTTAAATATGAGTTTTTCAATGTTTGAATGACATTTTGTCTCTCATTGATTCT
CATCTCATACCATCAACTACTACTAATGTGCAATCAATTGTCTTTTAAACAGAGTGAAA
GGAGATTATTTTCGATATATATGGCAGAGTTTGGTTCTGATGCTGAACGTAAAGAAAATGCA
GATAATTCTCTAGATGCATATAAGGTTGCAATGGAATGGCAGAGAATAGTTTAGCACCC
ACCAATATGGTTAGACTTGGATTGGCTTTAAATTTCTCGATATTCAATTATGAGATCCAT
AAATCTATTGAAAGCGCATGTAAATTGGTTAAGAAAGCTTACGATGAAGCAATCACTGAA
CTCGATGGCCTTGACAAGAATATATGCGAAGAGAGCATGTATATCATAGAGATGCTTAAA
TACAATCTTTCTACSTGGACTTCAGGCGATGGTAATGGTAATAAGACAGACGGTTAG
>G2123 Amino Acid Sequence (domain in AA coordinates: 99-109)
MRKVCELDIELSEERDLTTGYKNVMEAKRVSLRVISSIEKMEDSKGNDQNVKLIKQQ
EMVKYEFNVNCDILSLIDSHLIPSTTTNVESIVLFNRVKGDFRYMAEFGSDAERKENA
DNSLIDAYKVAMEAENSLAPTNMVRGLALNFSIFNYEIHKSIESACKLVKKAYDEAITE
LDGLDKNICEESMYIIEMLKYNLSTWTSBGDNGNKTDG*
>G2138 (27..512)
GGAACCCTAATTTCCGCAAAATTCATATGAAGCGTATTATCAGAATCTCATTACCGACG
CAGAAGCCACCGATTCTTCTAGCGACGAAGACACGAGGAGCGTGGAGGAGCATCCCGA
CTCGGCGCGGTGGGAAACGCCTCGTTAAAGAGATCGTAATCGATCCTTCCGATTCCGCCG

ATAAACTCGATGTCGCAAAACACGGTTCAAATCAGGATCCCGCGGAATTTCTCAAGA
CGGCGAAAACGGAGAAGAAATATCGTGGAGTGAGGCAGAGCCGTGGGGGAAGTGGGTGG
CGGAGATCAGATGTGGAAGAGGAGCTTGTAAGGACGACGTGATCGTCTCTGGCTGGGTG
CTTTTAACACTGCTGAGGAAGCTGCTCTAGCTTATGATAACGCTTCAATTAAGCTGATTG
GACCTCACGCGCCGACCAATTTTGGTTTGCCGGCGGAGAATCAAGAGGATAAGACGGTGA
TTGGAGCTTCTGAGGTTGCTAGAGGCGCGTGAAGTGGGGTTGGTAATTTAGTTGTTAGC
>G2138 Amino Acid Sequence (domain in AA coordinates: TBD)
MKRIIRISFTDAEATDSSSEDETEERGASQTRRRGKRLVKEIVIDPSDSADKLDVCKTR
FKIRIPAEFLKTAKTEKKYRGVRQRPWGKQWVAEIRCGRGACKGRRDRRLWLGTFNTABEAA
LAYDNASIKLIGPHAPTNFGLPAENQEDKTVIGASEVARGA*

>G2139 (40..663)
CCTACAAGAAATCAAACACTAGTTCTGGTTTCTGCAACATGTCATCTACGAAGCAAGCA
AAGGGAAGAAAACAAAGGGGAAGCAAAGATCGAGATGAAGAAGGTGGAGAAGTATGGA
GATAGGATGATTACGTTCTCAAACGTAAAACCGGAATTTTAAAGAAATGAACGAGCTC
GTAGCAATGTGTGACGTTGAAGTGGCTTTCTTGATTTTCTCTCAACCCAAGAAGCCCTAT
ACATTTCGCACATCCGTCTATGAAGAAAGTGGCTGACCGGTTAAAGAACCCCTTCGAGACAA
GAACCATTAGAGAGAGACGATACCAGACCCCTCGTCAAGCTTATAAGAAACGAAGGCTC
CAGGACCTCGTAAAAAAATGGAGGCGCTCGAAGAGGAGCTGCGATGGATCTAGAGAAG
TTGAAAACGTGTTGAAGGAATCGAGAAATGAAAAGAAGTTAGATAAAATGTGGTGAACTTT
CCTTCGGAAGGTTTGAAGCGGAAGGAGCTGCAGCAAAGGTACCAAGCGATGCTCGAGTTA
CGTGATAACTTATGCGACAATATGGCTCACTTACGATTGGGAAAAGACTGTGGTGGTTCA
TCTTCTGTTTCGTGTTGGGACGTCGAGTTTCTGGTGGTGTTCGTCTGTTTCGATCGTGAAGCA
TGATCATACATATTCATACTTGATGATTTAAATTTCTTTGTATTTGAACGCTGATTTTA
ATACTGCATGTATCCATTGACGAAGCTCAATCGTCTCGAGTATATCTCTATTATCTAAC
AGTATTGAGAAAAAAGGAGTTTCAGTAAAAAAGGAGTTTTCAGTAAAAAAGGAGTTTTCAGT
>G2139 Amino Acid Sequence (conserved domain in AA coordinates: 14-69)
MSSTKQAKGRKTKGKQKQIEMKKVENYGDRLMITFSKRKTGIFKKMNLVAMCDVEVAFLIF
SQPKPYTFAHPSMKKVADRLKNPSRQEPLEDRDTRPLVEAYKKRRLHDLVKMEALEEE
LAMDLKLLKLLKESRNEKKLDKMWNFPSSEGLSAKELQRYQAMLELRDNLCDNMAHLRL
GKDCGSSSVRVGRRVSGGVRLFDREA*

>G2343 (1..1113)
ATGGGTCATCACTCATGCTGCAACCAGCAAAAGGTGAAGAGAGGGCTTTGGTCACCGGAA
GAAGATGAGAAGCTTATTAGATATATCACAACCTCATGGCTATGGATGTTGGAGTGAAGTC
CCTGAAAAAGCAGGGCTTCAAAGATGTGGAAAAAGTTGTAGATTGCGATGGATAAACTAT
CTTCGACCTGATATCAGGAGAGGAAGGTTCTCTCCAGAAGAAGAGAAATTGATCATAAGC
CTTCATGGAGTTGTGGGAAACAGGTGGGCTCATATAGCTAGTCATTTACCGGGAAGAACA
GATAACGAGATTAATAAACTATTGGAATTCATGGATTAAGAAAAAGATACGAAAACCGCAC
CATCATTACAGTCGTCATCAACCGTCAGTAACACTACTGTGACATTGAATGCGGACACTACA
TCGATTGCCACTACCATCGAGGCTCTACCACCACAACATCGACTATCGATAACTTACAT
TTTGACGGTTTCACTGATTCTCTAACCAATTAATTTTACCACATGATCAAGAACTAAT
ATAAGATTCAAGAACTTTTTTCTCCATAAACCTCCTCTCTTCATGGTAGACACAACA
CTTCCTATCCTAGAAGGAATGTTCTCTGAAAACATCATCACAACAATAACAAGAACAAAT
GATCATGATGACACGCAAGAGAGGAAGAGAAAAATGTTTGTGAACAAGCATTTCTAACA
ACTAACACGGAAGAATGGGATATGAATCTTCGTGACGAAGAGCCGTTTCAAGTTCCTACA
CTGGCGTCACATGTGTTCAACAACCTTTCCAATTCAAATATTGACACGGTTATAAGTTAT
AATCTACCGCGCTAATAGAGGGAATGTGCGATAACATCGTCCATAATGAAAACAGCAAT
GTCCAAGATGGGAGAAATGGCGTCCACATTGCAATGTTTAAAGAGGCAAGAACTAAGCTAT
GATCAATGGGACGAFTCACAACAATGCTCTAATTTTCTTTTGGGACAACCTTAATATA
AACGTGGAAGGTTTCTCTTGTGGAACCAAGACCCATCAATGAATTTGGGATCATCT
GCCTTATCTTCTTCTTCCCTTCTCGTTTTTA

>G2343 Amino Acid Sequence (domain in AA coordinates: 14-116)
MGHSCCNQKVKRGLWSPEEDEKLIRYITTHGYGCWSEVPEKAGLQRCGKSCRLRWINY
LRPDIRRRFSPPEEKLIISLHGTVGNRWAHIAHLPTDNEIKNYWNSWIKKIRKPH
HHYSRHQPSVTTVTLNADTTSIATTIEASTTTTSTIDNLHFDGFTDSPNQLNFTNDQETN
IKIQETFFSHKPPLFMVDTTLPILEGMFSENIITNNKNNDHDDTQRGGRENVCEQAFIT
TNTEEDWMNLRQQEPFQVPTLASHVFNNSSNSNIDTVISYNLPALIEGNVDNI VHNENSN
VQDGEMASTFECLKRQELSYDQWDSQQCSNFFFDNLNINVEGSSLVGNQDPSMNLGSS

ALSSSFPSF*

>G265 (280..1317)

CTTTGGTCTTGGAAGCCAAATCAAACCTTTCCTTCAATCCTCAAATTTTCGAAAATTTTC
TCTTTTGCTTTACGTTCTCTCAATTCTATTTGTAAGAAAGTTTGTTCCTTTAATCAATC
AAATCAAAGAGACTTTTGAAGATTGTTTCCCAATTTGCGTCAATCGGGATCGAGTCAAAT
CTGAAATCTTCTCCACTCATCATCTGACTATAAGACTTAATCAAGGGACTTTTGTTCGG
GTTTGGTTTAAACGTCTTGGATTGGAAGTGGTTAAGGTATGGATGAAAATAATGGAGGT
TCAAGCTCACTTCCACCTTTCCTTACTAAAACATATGAAATGGTTGATGATTCTTCTTCT
GACTCGGTCTGTGCTTGGAGCGAAAACAACAAAAGCTTCATCGTCAAGAATCCAGCAGAG
TTTTCAAGAGACCTTCTTCCGAGATTCTTCAAGCATAAGAATTCTCAAGTTTCATCCGT
CAGCTTAATACATATGGTTTTTCGAAAAGTAGATCCTGAGAAATGGGAATTCTTGAATGAT
GATTTTGTAGAGGTCGACCTTACCTTATGAAGAACATTTCATAGACGAAAACCGGTTTCAT
AGCCACTCGTTAGTGAATCTACAAGCGCAAAATCCTTTGACGGAATCAGAAAACGCGAGC
ATGGAGGATCAGATAGAAAGACTGAAAAATGAGAAAGAAGGCCCTTCTTGGCGAGTTACAG
AACCAAGAGCAAGAACGGAAGAGTTTGAGCTGCAAGTAACGACATTGAAAGATCGGTTA
CAACATATGGAACAACATCAGAAATCAATAGTGGCATATGTTTACAGGTTTGGGAAAA
CCAGGACTTTCCTAAACCTCGAAAACCATGAGAGAAGAAAAAGAAGATTTCAAGAGAAC
TCTCTTCCCTCCAAGCAGTTTACACATAGAACAGGTCGAAAAGTTAGAATCTTCGCTAACG
TTTTGGGAGAATCTTGTATCGGAATCATGCGAGAAGAGCGGTTTGCAGTCATCAAGCATG
GATCATGATGCAGCTGAGTCAAGTCTAAGTATTGGCGATACAGACCCAAATCATCGAAG
ATTGATATGAATCAGAGCCGCCCTTACCGTTACTGCGCTGCTCCAAAAACAGGCGTT
AACGATGACTTTTGGGAACAATGTTTGACAGAGAACCCTGGATCAACCGAGCAACAAGAA
GTTCACTCAGAGAGAAGAGATGTCGGTAATGATAATAATGGTAATAAGATTGGAAATCAA
AGGACGTATTGGTGAATTCAGGGAATGTAAATAACATTACAGAGAAAGCTTCTTGACAT
GAATGAGGTTTTTTGAAAATAGTTTCTTTTGGTTCCACTGAGATTATTGTATGTGTTCA
TTATTTATTACTCTGTTTCTGTAAAAACAATCTCTCTATTGTTTGGGCAGGAGTGACA
TAAATGCATATGCAGAATTGGTTTCAAAA

>G265 Amino Acid Sequence (domain in AA coordinates: 11-105)
MDENNGSSSLPPLTKTYEMVDDSSDSVVAWSNNKSFIVKNPAEFSRDLLPRFFKHK
NFSSFIQRLNTYGRKVDPEKWEFLNDDFVRGRPYLMKNIHRRKPVHSHSLVNLQAQNPL
TESERRSMEDQIERLKNEKEGLLAELQNEQERKEFELQVTTLKDRLQHMEQHOKSIVAY
VSQVLGKPGLSLNLNHHERRKRFRQENSLPPSSSHIEQVEKLESSLTFWENLVSECEKS
GLQSSSMHDHAESSLSIGDTRPKSSKIDMNSPPVTVTAPAPKTGVNDDFWEQCLTENP
GSTEQQEVQSERRDVGNDDNNGNKIGNQRTYWNNSGNVNNITEKAS*

>G2792 (1..960)

ATGGATCATCATCATACATAGCATCAAGAAATTCATCAACAACATCAGAATTACCATCA
TTCGAGCCAGCGTGCCATAACGGTAATGGTAACGGTTGGATCTATGACCCAAATCAAGTT
AGGTACGATCAAGTAGTGACCAACGGCTGTCAAAGTTGACGGATCTTGTAGGCAAGCAC
TGGTCAATTGCAACACCGAATAATCCCGACATGAACCATAACCTTCATCATCACTTCGAT
CATGATCATTCTCAAACGACGACATTTCTATGTACAGACAAGCCTTGAGAGGTGAAAAAT
GAGGAAGATCTTTGTTACAATAATGGCTCAAGTGGTGGTGGTTCCTTGTTCATGATCCT
ATAGAAAGTTCTAGAAGTTTCTTGATATAAGGTTAAGTAGGCCATTAAACGGATATTAAT
CCGTCAATTAAAGCCATGCTTTAAGGCCTTAAACGTATCCGAGTTTAAACAAGAAAGACAT
CAAACGGCATCTCTGGCAGCAGTGAGACTGGGAACAACAAACGCTGGAAAAAAGAAGAGA
TGTGAAGAAATTTCCGATGAGGTTTCAAAGAAGGCCAAGTGCAAGTGGGGCTCTACACTT
TCGCCAGAGAAGGAATACCCAAAGCCAACTTCGAGACAAGATCAGGACTCTACAGCAA
ATTGTGTCTCCCTTTGGAAAGACTGATACTGCTTCTGTGCTTCAAGAGGCCATCACTTAC
ATAAATTTTATCAAGAGCAAGTTAAGCTGCTAAGCACTCCTTATATGAAGAATTCATCA
ATGAAGGATCCATGGGGGGGATGGGACAGAGAAGATCACAACAAAGGGGACCGAAGCAT
CTTGATCTAAGGAGTAGAGGGCTTTGTTGGTTCTTATTTTATATACCCCAATCGCATAC
CGCGATAACAGTGCAACTGACTACTGGAATCCACGTATAGAGGTTCTTTGTATCGTTAG
>G2792 Amino Acid Sequence (domain in AA coordinates: 190-258)
MDHHHHIASRNSSTSELSPFEPACHNGNGNGWIYDPNQVRYDQSSDQRLSKLTDLVGKH
WSIAPPNNPDMNHNLHHFHDHDSQNDISMYRQALEVKNEEDLCYNNGSSGGGSLFHD
IESRSFLDIRLSRPLTDINPSFKPCFKALNVSEFNKKEHQASLAAVRLGTTNAGKKKR
CEEISDEVSKKAKCSEGSTLSPEKELPKAKLRDKITLQQIVSPFGKTDASVLQEAITY
INFYQEQVKLLSTPYMKNSSMKDPWGGWDREDHNRKGPKHLDLRSRGLCLVPISYTPAIY

>G2830 (1..903)

>G2830 Amino Acid Sequence (domain in AA coordinates:245-266)
MSSIPNRFNIYGGDTTNHRESLPIEMNHNSRMVRSMFITSDRMNHRDLFSSPPSFSSYQN
SHISSSSVGFNNSHMTYHMLKRNYSVSRADYFSTKDHSHFTQVSFTQTITNKYTTIVPS
NIFDTVHYDIGRVKRAIDFRNIWNPKSHLPKKKFNRQCEILNPPLNIVFPHQDSADRQHL
DIIFSSSKHNHVFQDGRSLKKISEPTNLF EKSN SYDSQEDEKIDAYQYDGRTHSLPYTKY
GPYTCPRCNGVFDTSQKFAAHMLSHYNNETDKERDQRFRRNKKRYRKFMDSLKISKQKI
*

>G286 (94..2454)

TGCAATTTCCTCTCGACCAAAACCTAATTTACAGGTTTGGGGTTTTCTTCCTTCTACTGTCT
AATTTTGATGAACTTGTGATTCACTGATTAGAATGAATGCTAATGAGCAAACCTCGATCC
GCCAATGGCATTGGCAATGGCAATGGTGAGTCTATTCCCGGGATTCCAGATGACTTACGG
TGCAAGAGATCGGATGGTAAACAGTGGAGATGCATGCAATGTCCATGGCTGATAAGACT
GTTTGTGAGAAGCACTAATCCAAGCAAGAAGCGGGCTAATTTCTGCTTTCAGGGCG
AACCAGAGAAAGCGCAAAAGGCGATCATCGTTAGGCGAAACAGATACGTATTTCCGAAGG
AAGATGGATGATTTCGAGTTACCACTCACCAGCATTGACCACTATAATAACGGTCTTGCC
TCTGCTTCCAAGAGTAATGGTAGACTAGAGAAGAGACATAATAAAAGCCTGATGCGGTAC
TCGCCCGAGACACCGATGATGAGGAGTTTCTCTCCACGTGTTGCAGTGGATTGAATGAT
GACTTGGGTAGAGATGTTGTAATGTTTGAAGAGGGCTACAGATCTTATAGGACACCACCA
TCTGTTGCTGTTATGGATCCGACAGCAACAGATCACACAAAGCACCAGCTTATGGAA
TACTCAGACGACAGCACAGATGTGCTGCAAGAGTCTTTGGGGGAAATCTGCCATCAATGC
CAGAGAAAAGATAGAGAGAGAATCATTTCTTGCCCTCAAATGCAATCAAAGAGCCTTCTGC
CACAATTGTCTATCGGCAAGGTACTCGGAGATATCACTTGAAGAAGTCGAGAAAGTTTGC
CCTGCATGTGCTGGCTGTGTGATTGCAAATCTTGCTGCGTTAGATAATACAATAAAG
GTTCCGATCCGGGAAATACCCGTTTTGGACAAGTTGCAAGTATCTTTATCGTCTATTATCA
AGTGCTCTACCAGTCATAAAGCAGATCCATCTTGAACAATGTATGGAAGTTGAACTAGAG
CAGAGGCTTCTTGAAGTTGAGATTGATCTTGTCAGGCGAAGATTGAAAGCAGATGAGCAG
ATGTGCTGCAACGTGTGTGCGATACCAGTTGTTGACTACTACCGTCACTGTCCGAACTGC
TCATATGACCTTTGCCGTGAGATGCTGTCAAGATCTACGGGAAGAGTCTTCAGTGACGATT
AGTGGGACTAACCAAAACGTACAAGATAGAAAAGGAGCTCCCAAACTAAAACCTAACTTT
TCATACAAGTTTCTTGAGTGGAAGCCAACGGTGATGGGAGCATCCCTTGCCCTCCTAAG
GAGTATGGAGGCTGCGGTTACATCTTTGAATCTTGCCCGCATTTCAAGATGAATTGG
GTTGCAAAGTATGTGTAAGAACTCTGAGGAGATTGTTAGTGCGTCAAATATCTGATCTT
CTGAAACCTGATATGTGTGATTCAAGATTCTGCAAATTTGCTGAGAGAGAAGAGAGCGGT
GACAACCTACGTGTACAGCCCGTCGCTTGAAACGATTAAAACCTGATGGAGTAGCTAAGTTT
GAGCAACAATGGGCAGAGGGTCGGCTTGTTACTGTGAAAATGGTACTTGATGACTCATCT
TGCTCTAGATGGGATCCTGAGACTAATTTGGAGGGATATAGACGAGCTTTCCGACGAGAAA
CTGAGAGAACATGATCCATTCTTGAAGGCCATTAAATGCTTGATGGTGTAGAGGTGAT
GTAAGACTTGGGAGTTTACAAGAGCATATAAAGATGGAAGAACCAAGAGACAGGTCTT
CCGCTATTGTGGAAGTTAAAGGACTGGCCGAGCCCAAGTGCTTCCGAGGAGTTCAATTTTC

TACCAAAGACCTGAGTTTATCAGAAGTTTTCCGTTTCTCGAGTACATTTCATCCCCGGTTA
GGCCTTCTGAATGTTGCAGCCAAGTTACCTCATTACTCGCTCCAAAACGATTTCAGGTCCA
AAGATTTATGTGCTTGTGGGACGTACCAAGAAATCAGTGCTGGCGATTTCATTGACTGGT
ATTCACTACAACATGCGTGACATGGTATACCTATTGGTGACACGTCTGAAGAAACAACA
TTCCGAAAGGGTGAGAAAAACAAACCTGTTCCAGAGGAACCTGACCAGAAGATGAGCGAA
AATGAGTCACTTCTTAGCCCTGAGCAGAAATTAAGGGACGGAGAGTTACATGATCTATCA
CTTGGTGAAGCCAGTATGGAGAAGAATGAACCTGAGTTGGCGTTGACTGTGAATCCAGAG
AACTTAACGGAAAAACGGTGACAACATGGAATCTTCTGCACATCTTCATGTGCAGGAGGA
GCCCAGTGGGATGTCTTTTCGACGCCAAGACGTCCCAAAGTTGTCCGGGTATTTGCAGAGA
ACATTCCAGAAGCCTGATAATATCCAGACTGATTTTGTAAAGCCGTACCTGCTAATTCAAA
TAAATGAAGTGTGTAAAGTCTTGTATGTGGAATGATTGAGTTTCTAGTTTGTACTCT
GGTTTCAGGTGTACGCCCCGTTGTATGAAGGATGTCTTTAAATGAACACCACAAGAGAC
AACTAAGAGACGAGTTTGGAGTTGAGCCATGGACATTTGAGCAACATCGTGGTGAGGCTA
TCTTCATTCCGGCTGGATGTCCGTTCCAAATCACTAATCTTCAGTCCAATATTCAGGTGG
CACTTGACTTCTTGTGCCCTGAAAGCGTTGGAGAGTCAGCAAGACTAGCTGAAGAAATCC
GGTGTTTACCAAACGACCACGAGGCAAAACTTCAGATTCTAGAGATTGGAAGATATCAT
TATACGCAGCTAGCTCAGCCATTAAAGAGGTTTCAAGAACTGGTCTTGGATCCAAAGTTTG
GAGCAGAGCTTGGATTTGAAGACTCTAACTTAACCAAAGCAGTCTCTCACAACCTAGACG
AGGCAACCAAGCGGCC

>G286 Amino Acid Sequence (domain in AA coordinates: TBD)
MNAEQTRSANGIGNNGESIPGIPDDLRCRSDGKQWRCTAMSMADKTVEKHYIQAKK
RAANSAFRANQKAKRRSSLGETDITYSEGKMDDFELPVTSIDHYNGLASASKSNGRLEK
RHNKSLMRYSPETPMRFSFPRVAVDLNDDLGRDVVMFEEGYRSYRTPPSVAVMDPTRNR
SHQSTSPMEYSAASTDVSAESLGEICHQCQQRDRERIISCLKCNQRAFCHNCLSARYSEI
SLEEVEKVCPRGLCDCKSLRSDNTIKVRIREIPVLDKLYRLLSAVLPVIKQIHL
EQCMEVELEKRLLEVEIDLVRARLKADEQMCCNVCRIPVVDYYRHCPNCSYDLCLRCCQD
LREESSVTISGTNQNVQDRKGAPKLKLNFSYKFPWEANGDGSIPCPPEYGGCGSHSLN
LARIKMNWVAKLVKNAAEIVSGCKLSDLLNPDMDCSRFCFAEREESGDNYVYSPSLET
IKTDGVAKFEQOWAEGRLVTVMVLDDSSCSRWDPETIWRDIDELSDEKLREHDPFLKAI
NCLDGLLEVDRLGEFTRAYKDGNQETGLPLLWKLKDWSPSPASEEFIYQRPFIKRSFP
FLEYIHPRLGLLNVAALKPHYSLNDSGPKIYVSCGTYQEISAGDSLTIHYNMRDMVYL
LVHTSEETTFEVRVKTTPVPEEPDQKMSSENLSPQKLRDGLHDLGLGEASMEKNEP
ELALTVPENLNTENGDNMESSCTSSCAGGAQWDVFRQDVPKLSGYLQRTFQKPDNIQTD
FVSRTC*

>G291 (124..1197)
CAAGAACCCAAAGATCTCTCTCTATTGTTTGCCTTCTTCTTTCTTTCTGACTCAAACCC
TCAAATCAATTCTCGCGATTAAGCAAAACCTAGATTATTCTACTCTTCGAAGTCGATT
TCAATGGAGGTTCTCTCGTCAGCCATCGCGAGGAAGACATGGGAGCTAGAGAACAACATT
CTCCAGTGAACCAACCGATTGAGCTCCGACAGTATATTCCACTACGACGACGCTTCA
CAAGCCAAAATCCAGCAGGAGAAGCCATGGGCCTCCGATCCTAACTACTTCAAGCGCGTT
CACATCTCAGCCCTTGCTCTTCTCAAGATGGTGGTTACGCTCGCTCCGGTGGCACAATC
GAGATCATGGGTCTTATGCAGGGTAAACCGAGGGTGATACAATCATCGTTATGGATGCT
TTTGCTTTGCCTGTTGAAGGTACTGAGACTAGGGTTAATGCTCAGTCTGATGCCATGAG
TATATGGTTGAATACTCTCAGACCAAGCAAGCTGGCTGGGAGGTTGGAGAACGTTGTTGA
TGGTATCACTCTCACCCTGGGTATGGATGTTGGCTCTCGGGTATTGATGTTTCGACACAG
ATGCTTAACCAACAGTATCAGGAGCCATTCTTAGCTGTTGTTATTGATCCAACAAGGACT
GTTTCGGCTGGTAAGGTTGAGATTGGGGCATTGAGAACATATCCAGAGGGACATAAGATC
TCGGATGATCATGTTCTGAGTATCAGACTATCCCTCTTAACAAGATTGAGGACTTTGGT
GTACATTGCAAAACAGTACTACTCATTTGGACATCACTTATTTCAAGTCATCTCTCGATAGT
CACCTTCTGGATCTCTTTTGAACAAGTACTGGGTGAACACTCTTCTTCTTCCCACTG
TTGGGCAATGGAGACTATGTTGCCGGGCAAATATCAGACTTGGCTGAGAAGCTCGAGCAA
GCGGAGAGTCAGCTCGTAACCTCCGGTATGGAGGAATTGCGCCAGCCGGTCACCAAAGG
AGGAAGAGGATGAGCCTCAACTCGCGAAGATAACTCGGGATAGTGCAAAGATAACTGTC
GAGCAGGTCCATGGACTAATGTACAGGTTATCAAAGACATCTTGTTCATTCCGCTCGT
CAGTCCAAGAAGTCTGTGACGACTCATCAGATCCAGAGCCCATGATTACATCGTGAAGT
TGGTCTATTCTTTTCTTTTGGCTGCGGAAATGACTATCGGTTTGACCCGGTTTATGA
GGCAATGCCCATTTGTTCCCTATATCTCTAGTGTAGTATCTGCTTCAGACAAAGATCTTTG

GGTTATTAAATGACATTAACATAAAAAA
>G291 Amino Acid Sequence (domain in AA coordinates: 132-160)
MEGSSSAIARKTWELENNILPVEPTDSASDSIFHYDDASQAKIQQEKPWASDPNYFKRVH
ISALALLKMMVHARSGGTIEIMGLMQGKTEGDTIIVMDAFALPVEGTETRVNAQSDAYEY
MVEYSQTSKLAGRLENVVGWYHSHPGYGCWLSGIDVSTQMLNQYQEPFLAVVIDPRTV
SAGKVEIGAFRTYPEGHKISDDHVSEYQTIPLNKIEDFGVHCKQYYSLDITYFKSSLDSh
LLDLLWNKYWVNTLSSSPLLNGNDYVAGQISDLAEKLEQAESQLANSRYGGIAPAGHQRR
KEDEPQLAKITRDSAKITVEQVHGLMSQVIKDILFNSARQSKKSADSSDPEPMITS*

>G427 (49..1230)
TTTCCTCTCCGAAACAGAAATTCAAAAACAAATTCACACGAAAACGATGGCGTTTCAT
AACATCACTTTAATCATTTTACCGACCAACAACATCAGCCTCCTCCTCCGCCGCAA
CAGCAGCAGCAACAACATTTTCAAGAATCAGCACCCCCCTAATTGGCTCCTCCGCTCCGAC
AACAACTTCTCAATCTCCACACAGCTGCCACAGCCGCCGCTACAAGCTCCGATTCTCCT
TCTTCCGCCGCCGCTAACCAGTGGCTCTCACGATCCTCATCCTTCTCCAACGAGGCAAC
ACCGCAACAACAACAACAACGAAACATCCGGTGACGTATCGAAGACGTTCCCGCGCGGA
GAGGAGTCAATGATCGGAGAGAAGAAGGAGGCGGAGAGGTGGCAGAATGCGAGACACAAG
GCGGAGATACTGTCTCATCCACTATACGAGCAACTTTGTGCGCACACGTGGCGTGCCTG
AGGATCGCAACGCCGTGGATCAGCTTCCGAGGATAGACGCACAGCTTGCTCAGTCTCAA
AAGCTCGTGGCTAAGTACTCAACTTTAGAAGCCGCTCAAGGACTCCTCGCCGCGCATGAC
AAGGAGCTTGACCACTTCATGACGCATTATGTACTATTGCTTTGCTCTTCAAAGAACAA
CTGCAACAGCATGTTCTGTTCATGCAATGGAAGCTGTTATGGCCTGTTGGGAGATTGAA
CAGTCGCTTCAAAGTTTTACAGGAGTATCTCCTGGTGAAGGCACAGGAGCAACAATGTCT
GAGGATGAAGATGAGCAAGTAGAGAGTGATGCTCATTTGTTTGATGGAAGCTTAGATGGG
TTAGGGTTTGGTCTCTAGTTCCCACTGAGAGCGAGAGATCTTTGATGGAACGAGTCAGA
CAAGAACTCAAACATGAACCTCAAGCAGGGTTACAAGGAGAAAATTGTGGACATAAGAGAG
GAGATACTGAGGAAGAGAAGAGCTGGAAAATTACCAGGAGACACCACCTCTGTTCTCAA
TCATGGTGGCAATCTCATTCTAAGTGGCCTTACCCTACTGAGGAAGATAAGGCGAGGTTG
GTGCAGGAGACGGGTTTGCAGCTCAAACAGATAAACAATTGGTTCATCAATCAAAGAAAG
AGGAATTGGCATAGCAATCCATCTTCTTCTACCGTCTCAAAGAATAAACGCCGAAGCAAT
GCAGGTGAAAACAGCGGAAGAGACCGTTGAGATCAAGCTTGATGTAGAGATCCAAAAGC
TTTATAGAAAGGTGGAGGCATGAAGACAAAGAATTCTTACACAACAACGTAGGACGTAA
TTTTGTGCCAGTACATGGTATGGCTTTTCAATTTGGTAATGATTAGGGCCACACAAAATT
AAACCCCAAAGCATGATTTGTAATATGAGGTTTTAGATGGACTTTATGATAGGATCGTCA
GTCTTCACTGCCATCTCCATTCTCCACCATCAATCCATCATTATATCTTGTGAAAAAAA
A

>G427 Amino Acid Sequence (domain in AA coordinates: 307-370)
MAFHNNHFNHFTDQQQHQP PPPPQQQQQHFQESAPPNWLRLSDNNFLNLHTAATAAATS
SDSPSSAAANQWLSRSSSFLQRGNTANNNNNETSGDVIEDVPGGEESMIGEKKEAERWQN
ARHKAELSHPLYEQLLSAHVACLRIATPVDQLPRIDAQLAQSONNVVAKYSTLEAAQGLL
AGDDKELDHFMTHYVLLLCSPKEQLQQHVRVHAMEAVMACWEIEQSLQSFTGVSPGEGTG
ATMSEDEDEQVESDAHLFDGSLDGLGFGPLVPTESERSLMERVRQELKHELKQGYKEKIV
DIREILRKRRAGKLPDGTTSVLKSWWQSHSKWPYPTEEDKARLVQETGLQLKQINNWF I
NQRKRWNHNSNPSSSTVSKNRRSNAGENSGRDR*

>G509 (122..1054)
CTTCCTCCTTTGCTAATAAACTTTTCTTTGAACCTTACACGCCTTGTGATATTACTCTC
TTAAATATATATTTTCGTACATTAACACAGACATATATAAGCTAAAGATTTCTTCAGT
AATGGGTTTGAAAGATATTGGGTCCAAATTGCCACCGGGTTTCGATTTTCATCCAAGTGA
TGAAGAGTTGGTTTGTCTTATCTTTGCAACAAGATTAGGGCCAAATCTGATCATGGTGA
TGTTGATGATGATGATGATGATGTTGATGAAGCTTTGAAGGGTCTACTGATCTTGTGGA
GATTGACTTGATATCTGTGAGCCATGGGAGCTTCCTGATGTGGCAAAGTTAAACGCAAA
GGAATGGTACTTCTTCAGTTTCCGTGATCGAAAGTATGCTACTGGATATCGCACGAACAG
AGCGACAGTAAGCGGATACTGGAAGCAACAGGAAAAGATCGAACGGTGATGGATCCACG
TACAAGGCAATTGGTAGGGATGAGAAAAACTAGTGTTCTACAGAAACAGAGCACCAAA
TGGGATCAAAACTACTTGGATCATGCACGAGTTCCGTCTTGAAGTGTCTTAACATCCCACA
TAAGGAAGACTGGGTCTTGTGCAGAGTGTTCACAAAGGCAGAGACTCATCGCTACAAGA
CAATAATTATTATAACAATGATAATCAGACGCAAGGCTTGAAGTTAATGACGCTCCGGA
TCTTAATTACAACAATCAGTTGCCACCTTTGCTATCATCCCTCCTCATATCATCAACA

TGAGAAGATGAAAATCCAAGTTTGTGATCAGTGGGAGCAGCTAATGAAGCAGCCTTCAAG
GACCACCGGCCACCCCTATCATCACCATTGTCATCATCAAACCATAGCATGTGGTTGGGA
GCAGATGATGATCGGTTGCTGTGTCATCACCTTCGAGTCATGGCCCTGATCAGAGTCCTT
TGCTAAATTTGCTTTACCGTCGACAATAACAACAGTGTCAACATCAGTGGTGATCATCAT
CAGAATTATGAGAAGATTTTGTGTCATCACTAGACATGACGAGTTTGGATCAGGACAAG
ACATGTATGGGATCATCATCGGATGGTGGTATGGTCTCTGATCTTCACATGGAATGTGGT
GGATTGAGTTTGTAGACCGAAAATATCCTCGCTTTCCAATGAACATAATTCAAGGGGTTT
GCCAATTTGTGATTCTGTAATTATACAAACATTTTATCTATAGATTTATCACATTATCA
AACATGTAAGTTGTGTGGCATTGTTGGGTATAGGGTTTGTGTGATTCTAGGTTTTTAGGACG
ATGTATGTTGTTATATTTAGCGTGTTTTATAGGATTTATTCTCATTTTAAATATATGAA
AACCATTACTATGAATACAATTAGTTTCTTTGTTGTAAATAATATTTTAGATTATCAA
AAAAAAAAAAAAAA

>G509 Amino Acid Sequence (domain in AA coordinates: 13-169)
MGLKDIGSKLPPGFRFHPSPDEELVCHYLCKIRAKSDHGDVDDDDDDVDEALKGSTDLVE
IDLHICEPWELPDVAKLNKAWYFFSFRDRKYATGYRINRATVSGYWKATGKDRTVMDPR
TRQLVGMKRKTLVFNRRAPNGIKTWMHEFRLECPNIPHKEDWVLCRVFNKGRDSSLQD
NNYNNNDNQTRLEVNDAPDLNYYNQLPPLLSSPPHNQHEKMKIQVCDQWEQLMKQPSR
TTGHPYHHCHHQTACGWEQMMIGSLSSPSSHGPDHESFAKFALPSTITTVSTSVVIII
RIMRRFCCHH*

>G519 (85..894)

CACAAAGATCCTCCGATTTCGAAGGTTTATAAAAACTCAAATCGAATCTTATCCACAAGA
AAACAACAGGTACTTTTCAAAAATGAAGGCGGAGTTGAATTTGCCGGCGGGATTCCGA
TTTCATCCGACGGACGAAGAGCTTGTCAAGTTCATCTTTGCCGGAGATGTGCGTCAGAA
CCGATTAACGTTCCGGTTATCGCAGAGATTGACTTGTACAAATCAATCCATGGGAGCTT
CCAGAAATGGCGTTGTACGGTGAGAAAGAATGGTACTTCTTCTCGCATAGAGACCGGAAA
TACCCAAACGGGTCGAGACCAAACCGGGCAGCTGGAACCGGTATTGGAAAGCGACTGGA
GCTGATAAACCGATCGGAAAACCGAAGACGTTAGGGATTAAGAAAGCACTCGTCTTCTAC
GCAGGAAAAGCTCCGAAAGGGATTAAACGAATTGGATTATGCACGAGTATCGTCTCGCT
AATGTCGATCGATCTGCTTCTACCAACAAGAACAACAACCTTAAGACTTGATGATTGGGTT
TTGTGTCGGATATACAATAAGAAAGGAACAATGGAGAAGTATTTACCGCGCGCGCTGAG
AAACCGACGGAAGATGAGTACGTCGGACTCAAGATGCTCAAGTCACGTGATTTACCCG
GACGTCACGTGTTCTGATAACTGGGAGGTTGAGAGTGAGCCCAATGGATTAATCTGGAA
GACGCGTTAGAGGCATTTAATGATGACACGTCATGTTTAGTTCATTGTTTGTGCAA
AATGACGCTTTGTTCTCAGTTTCAGTACCACTCCTCCGATTTCGTGATTCGTTTCAG
GACCGTTTCGAGCAGAAACCGTTCTTGAATTGGAATTTTGCTCCTCAAGGGTAAAAATAA
TCGGCAAAAAGTTGAAGCTTTTCAGAGTCTTCGATCACCAGCATTGTGTGCGATCCTGAC
CCGGAGACCAAGTCGGGTACATACGATTACATAATCGGGTTATTGAGATTTCACATTGG
ATTTCCGAGACTAACCAACTTAACGGATTCTGGGGTAATTGGGGGGTTTGCACAGGTGA
ATCACACTGAGTCAGCAAGTTTCGATTTTGTGTTTGTGTTTGTAAATGATTGATTAAATG
TCTAAAGATATACGAAGTAGATTGAGAAGCACTGTAAGCAATTGTGACCACCCGTTA
TGAATCATAAATATATTCAATGAAGCATGAGCTTATTTTAAAAA

>G519 Amino Acid Sequence (conserved domain in AA coordinates: 11-104)

MKAELNLPAGFRFHTDEELVKFYLCRRCASEPINVPVIAIDLYKFNPWELPEMALYGE
KEWYFFSHRDRKYPNGSRPNRAAGTGYWKATGADKPIGPKTLGIKKALVIFYAGKAPKGI
KTNWIMHEYRLANVDRSASTNKKNNLRLLDDWVLCRIYNKKGTMEKYLPAAEKPTKEMST
SDSRCSSHVISPVTCSNWEVESEPKWINLEDALEAFNDDTSMFSSIGLLQNDAFVPQF
QYQSSDFVDSFQDPFEQKPLNWNFAPQG*

>G561 (86..1168)

AATTTGTTTTTTTTCTTTTGTGGGTTCAATTGCAATTGTTTTCCCTGAGACTCAAGTTA
CTGTGTCATTACTCTGCATTGAGCAATGGGTAGCAACGAAGAAGGAAACCCACTAACAA
CTTGATAAGCCATCGCAAGCTGCTGCTCCTGAGCAGAGTAATGTTTCATGTGTATCATCA
TGACTGGGCTGCTATGCAGGCATATTATGGGCCTAGAGTTGGTATACCTCAATATTACAA
CTCAAATTTGGCGCTGGTCATGCTCCACCGCTTATATGTGGGCGTCTCCATCGCCAAT
GATGGCTCCTTATGGAGCACCATATCCACCATTTTGCCTCCTGGTGGAGTTTATGCTCA
TCCTGGTGTCAAATGGGCTCACAACCACAAGGTCCTGTTTCTCAATCAGCATCTGGAGT
TACAACCCCTTTGACCATGTATGCACAGCTAATTCAGCTGGAACCTCAGATCATGGGTT
CATGAAAAGCTGAAAGAGTTTCGATGGACTTGCAATGTCAATAAGCAATAACAAAGTTGG

GAGTGTCTGAACATAGCAGCAGTGAACATAGGAGTTCTCAGAGCTCCGAGAATGATGGCTC
TAGCAATGGTAGTGATGGTAATACTGGGGGAGAACAATCTAGGAGGAAAAGAAGGCA
ACAAAGATCACCAAGCACTGGTGAAAGACCCTCATCTCAAAACAGTCTGCCTCTTAGAGG
TGAAATGAGAAACCCGATGTGACTATGGGGACTCCTGTTATGCCACAGCAATGAGTTT
CCAAACTCTGCTGGCATGAACGGTGTGCCACAGCCATGGAATGAAAAAGAGGTTAAACG
AGAGAAGAGAAAAACAGTCAAACCGAGAATCTGCTAGGAGGTCAAGACTGAGGAAGCAGGC
TGAAACAGAACAACTATCTGTCAAAGTTGACGCATTAGTAGCTGAGAACATGTCTCTGAG
GTCTAAACTAGGCCAGCTAAACAATGAGTCTGAGAACTACGGCTGGAGAACGAAGCTAT
ATTGGATCAACTGAAAGCGCAAGCAACAGGGAAAACAGAGAACCTGATCTCTCGAGTTGA
TAAGAACAACCTCTGTATCAGGTAGCAAAACTGTGCAGCATCAACTGTTAAATGCAAGTCC
GATAACCGATCCTGTGCGGGCTAGCTGACCGTGCCGCAACAATGAGAACCCGATATTTC
TTCTTTTGGGTTGTGATTGTAACCTAAAAGGAGACTTTTTTGTTTTATTCTTAGATTTGT
AGCTCTCTGCATAGTGAGCATAAATTGATGTAATATGGTTTAAAGAGATTGGTGTCTCT
GGTGTGTGCTGCAACCACATAATTGGTGATAGATAGGTTTAGTTATATAAGCAAATGTAT
TAGAGATAAGGGGAGACATATTTGATGGTCTTT

>G561 Amino Acid Sequence (domain in AA coordinates: 248-308)
MGSNEEGNPTNNNDKPSQAAPEQSNVHVYHHDWAAMQAYYGPRVGIQYYNSNLAPGHA
PPPYMWASPSPMAPYGAOPYPPFCPPGGVYAHPGVQMGSPQGPVSQSASGVTTPLTIDA
PANSAGNSDHGFMKKLEFDGLAMSIENNKVGSAEHSSSEHRSSQSENDSNGSDGNT
TGGEQSRKRQRSPSTGERPSSQNSLPLRGENEKPDVTMGTPVMPTAMSFQNSAGMNG
VPQPWNEKEVKREKRQSNRESARRSRLKQAEQELSVKVDALVAENMSLRSLKGLQNN
ESEKLRLNEAILDQLKAQATGKTENLISRVDKNNSVSGSKTVQHQLLNASPIITDPVAAS

*

>G590 (102..1223)
TCGACAGACACTCTCCCTCTCTCCATGCCCATAAAATCTCAAAGACTGTTTAAAAA
AATGTTTTAGCTTTAACTGCTTTTTTTTGTGTTGTTGGTGTAATGATATCACAGAGAGAAG
AAAGAGAAGAGAAGAAGCAGAGAGTGATGGGAGATAAGAAATTGATTTCATCTTCTTCTT
CTTCTCGGTTTACGATACTCGTATCAATCATCATCTTCATCATCTCCGCTCTTCTTCCG
ACGAAATCTCTCAGTTTCTCCGGCATATTTTCGACCGTTCTTCTCCTTTACCTTCTTACT
ACTCCCGGCGCAGACTACAACGACGGCGTCTTTGATTGGTGTGCACGGGAGCGGTGACC
CACATGACGATAAATCTGAGAAGTCTCGTTTCTCATCATCCACCGTCAGATTCTGTGCTTA
TGTCGAAACGTGTGCGGAGATTTCTCTGAGGTTTAAATCGGCGGAGGATCAGGCTCAGCCG
CCGCGTGTTTTGGTTTCTCCGGTGGTGGTAATAATAACAACGTTCAAGGAAATAGCTCTG
GGACTCGAGTATCGTCTTCTTCCGTTGGAGCTAGTGGCAACGAGACAGATGAGTATGACT
GTGAAAGCGAGGAAGGAGGAGAAGCTGTAGTTGATGAAGCTCCCTCTTCCAAGTCAGGTC
CTTCTTCTCGTAGTTTATCTAAAAGATGCAGAGCTGCTGAAGTTCATAATCTCTCTGAGA
AGAGGAGGAGAAGTAGAATTAATGAAAAAATGAAAGCTTTACAAAGTCTCATCCCTAATT
CAAATAAGACGATAAGGTTTCAATGCTTGATGAAGCCATTGAGTATCTGAAACAGCTTC
AGCTCCAAGTTCAGATGTTGACTATGAGAAATGGAATAAACTTGCATCCTTTGTGTTTAC
CTGGAACCTACATTACACCCATTGCAACTCTCTCAGATTGACCCCTGAAGCAACCAATG
ATCCTCTGCTTAATCATACCAATCAGTTTGCTTCGACTTCTAATGCACCGGAAATGATCA
ATACTGTGGCTTCTTCATACGCTTTGGAACCTTCTATTTCGAGTCACTTTGGACCTTTCC
CTCTCCTTACTTCACCCGTGGAGATGAGTCGGGAAGGTGGGTAACTCATCCAAGGTTGA
ACATTGGTCAATCCAACGCAACATAACCGGGAACAAGCTCTGTTTGATGGACAACCTG
ACCTAAAAGATCGAATTACTTGAACAGTGTCCTCAACTTCGGGATCTCTATGTGTTCTGT
TTCTTAGAACGCAAGCCATAAAGCTGTCTGAC

>G590 Amino Acid Sequence (domain in AA coordinates: 202-254)
MISQREEREKKQVRMGDKKLISSSSSSVYDTRINHLHHPSSSDEISQFLRHIFDRS
SPLPSYSPATTTTTASLIGVHSGDPHADNSRSLVSHHPPSDSVLMSKRVGDFSEVLIG
GGSGSAAACFGFSGGGNNNVQGNSSGTRVSSSVGASGNETDEYDCSEEGGEAVVDEA
PSSKSGPSSRSSSKRCRAAEVHNLSEKRRRSRINEKMKALQSLIPNSNKTDKASMLDEAI
EYLKQLQLQVQMLTMRNGINLHPLCLPGTTLHPLQLSQIRPEEATNDPLLNHTNQFASTS
NAPEMINTVASSYALEPSIRSHFGPPFLTSPVEMSREGGLTHPRLNIGHSNANITGEQA
LFDGQPDLDKDRIT*

>G818 (65..1060)

GTATTTCTTACAATAAACGACCAAAAAGTTAATACAAGAAATAGAAACGGTGTAGGAAGC
TACTATGACGGCAATTCAAACGTCGTCGATATTGAATCTTCTTCTCTCTCGCTTTGTCA

AGAGACGGCAACGGAGACCGTCACCGTTGAAAGAGGCTCGTCTGATTTCATCTTCAAAGCC
AGACGACGTCGTTTTACTAATCAAGGAAGAGGAGGATGACGCCGTTAACTTGTCACCTGG
TTTTTGAAATTCACGAGATAGGTTTTAATAACACCGTTCTTGAGAAAGACGTTTGAGAT
CGTCGATGACAAAGTAACAGACCCGGTTGTATCATGGAGCCCGACCCGTAAAAGCTTTAT
CATTTGGGATTCTTACGAGTTCTCAGAGAATCTACTTCCCAAATACTTCAAGCACAAGAA
CTTCTCCAGTTTTATTTCGTAGCTTAACTCTTACGGTTTTAAAAAGGTCGATTGAGATAG
GTGGGAATTTGCTAACGAAGGGTTTCAAGGAGGGAAGAAACATTTGCTTAAGAACATCAA
GAGGAGAAGCAAAACACTAAATGTTGTAACAAGGAAGCGAGTACCACCACGACAGAGAC
TGAGGTTGAGTCATTGAAGGAGGAACAGAGTCCAATGAGATTGGAGATGTTGAAGCTGAA
ACAACAACAGAAGAATCTCAACATCAGATGGTCACTGTGCAGGAGAAGATCCACGGAGT
TGATACCGAACAACAGCATATGCTTAGTTTCTTTGCAAAGTTGGCTAAAGATCAAAGATT
TGTAAGAGAGCTGGTGAAGAAGAGAAAGATGAAAATACAGAGAGAGCTAGAAGCAGCTGA
ATTCGTGAAGAAGCTCAAGTTGCTTCAGGATCAAGAACTCAAAAGAACTTGTAGATGT
AGAAAGAGAATTTATGGCCATGGCTGCAACAGAACAAATCCCGAGCCTGACATTTTGGT
GAACAATCAAAGCGGGAATACGAGATGTCAGCTTAACTCAGAGGACCTACTTGTGACGG
TGGCTCAATGGATGTAAATGGGAGGATAGAGATAGAGTAGAGCAAAACCGGTAACATAGC
AATAGAGAAGGTACCAAATCCCAAGGCTTGAGATCCGAAT

>G818 Amino Acid Sequence (domain in AA coordinates: 70-162)
MTAIPNVVDIESSSSLCQETATETVTVVERGSSDSSSKPDDVLLIKEEEDDAVNLSLGF
WKLHEIGLITPFLRKTFEIVDDKVTDPVWSPTKRSLFIWDSYEPSENLLPKYFKHKNF
SSFIRQLNSYGFKKVDSRWEFANEGFQGGKKHLLKNIKRRSKNTKCCNKEASTTTTETE
VESLKEEQSPMRLEMLKLKQQQEESSQHQMVTVQEKIHGVDTEQQHMLSFFAKLAKDQRFV
ERLVKKRKMKIQRELEAAEFVKLKLQDQETQKNLLDVEREFMAMAATEHNPEPDILVN
NQSGNTRCQLNSEDLLVDGGSMDVNGRIEIE*

>G849 (218..2077)

AACTCGAGAATCTTTCATTTCTTTTAAATCTTAGAATCTCGAGTTTTTGTATAAATCGAT
TCTAATTTTCCTTTGTACATTGTTTTATATATACATAAAACACACAAATCGGGTATGGG
GGAATTTGGGTTTTAAGATAGCGTGATCTGTAATAATAAGTGGTTCGCGATCGTGATCAA
GAACTGGTGGCTGATAGTGATATGCATATTTAGAGATGGTGTTCAGAGAAAAGTTAGA
TTGCCCTTCCCGTGGGATTTGATTTTCCCAACATTTCCAGAGCTCCTCGTTCATGCAGGAG
GAAGGTTCTAAACAAGAGGATTGATCATGATGATGATAACACTCAGATCTGTGCAATTGA
CTTACTAGCTTTGGCTGGAAGATTCTACAGGAAAGCGAGAGTTCTCTGCGTCTTCTAA
TGCAATTTGAAGAAATTAAGCAAGAGAAAAGTAGAAAATTGCAAGACTATTAAATCTGAGTC
TTCTGACCAAGGAAACTCTGTGTCAAAGCCTACTTATGATATCTCTACTGAGAAGTGTGT
GGTGAACAGTTGTTTTTCATTTCCCGATAGTGACGCGCTTTTGGAGCGGACTCCGATGTC
TGATTACAAGAAGATTGATGTTTGTAGGATGTAGGGTGTGAAAAAAGAATGTAAATAA
TGGGTTTCGAGCAAGGAGAAGCAACCGATCGCGTGGGTGATGGAGGCTTAGTCACTGATAC
TTGCAACTTAGAGGATGCAACTGCGTTAGGTCTGCAGTTTCCGAAATCAGTCTGTGTGGG
TGGTGATTTAAAATCACCATCCACCTTGGATATGACCCCTAATGGTTCCATGCTAGACA
TGGGAACCATATAACCTAGGTAGAAAAGATGATGATGAAAAATCTATAGTTACCATAA
ACTTAGCAATAAAATTAAGTCGTATAGGTCTCCAACAATTCGAAGAATAAGAAAGTCCAT
GTCGTCCAAATACTGGAAACAAGTTCCAAAAGATTTTGGATACAGTAGAGCTGATGTGGG
TGTGAAGACTCTTTATCGCAAAAGAAAATCATGTTATGGTTACAACGCATGGCAGCGTGA
GATCATTTATAAGAGAAGAAGATCACCTGAAGAAGCTCGGTGTAACCTCTGATGGAGG
ACTCAGTAGTGGAAGTGTTCCTCAAGTTACCCAAGAAGGGAGATACAGTAAAGCTAAGCAT
TAAGTCCTTTAGGATTCCAGAGCTTTTTATTGAAGTTCCAGAACTGCAACAGTAGGATC
ACTAAAGAGGACTGTGATGGAGGCTGTCACTGTTTTACTCAGCGGAGGAATACGTGTTGG
GGTGTTAATGCATGGAAGAAGGTTAGAGATGAAAGGAAAATCTGTCCAGACTGGGAT
CTCATGTGATGAAAATCTAGACAACCTTGGGTTACCTTGGAGCCTAGTCCCAGCAAAGT
TCCCTACCTTTGTGTTCTGAAGATCCTGCTGTGCCAACCGACCCTACAAGTTTGTCTGA
ACGGTCTGCGGCTCTCCTATGCTAGATTCTGGAATCCACATGCAGATGACGTGATTGA
TTCAAGAAATATTGTGGACAGTAACCTCGAATTAGTTCATATCAGGGTGACATATCTGT
TGATGAACCTTCATCAGATTCAAAGAGCTTGTCCCACTTCCAGAGTTGGAAGTCAAGGC
GCTTGCCATAGTTCCGTTGAACCAGAAACCTAAGCGTACTGAGCTAGCCAGAGGAGAAC
TAGGAGACCTTCTCTGTGACAGAGGTAGAAGCTCTGTACAAGCAGTTGAGGAACTCGG
GACTGGAAGATGGCGTGATGTAAAATTGCGTGCTTTTCGAGGATGCAGATCATCGGACTTA
CGTGGACTTGAAGGACAAATGGAAGACGCTAGTTCACACAGCAAGTATATCCCCACAGCA

ACGAAGAGGAGAGCCGGTGCCACAAGAACTGCTAGACAGAGTCTTGAGGGCATAACGGGTA
TTGGTCGCAGCACCAAGGAAAACATCAGGCGAGAGGAGCGTCCAAAGATCCAGACATGAA
CAGAGGTGGAGCTTTTGAATCAGGTGTTTCAGTGTAAAAAAGGAGGTACGCATTGGTGGG
TGGGTGTACAGAAGCAAACAACAATAAATGGACAACCTCAATTCTGCAAAGTTTAATT
GTCTTTATTCTCGTTCTTTTTTTTTTTCTCCTACATACACTTTTTTTTTTCT

>G849 Amino Acid Sequence (domain in AA coordinates: 324-413, 504-583)

MVFKRKLDCLSVGFDFPNIPRPRSCRKVLNKRIDHDDNTQICAIDLLALAGKILQES
ESSSASSNAFEEIKQKVENCKTIKSESSDQGNVSKPTYDISTEKCVVNSCFSPDSDG
VLERTPMSDYKKIHGLMDVGCENKNVNNNGFEQGEATDRVGDGGLVTDTCNLEDATALGLQ
FPKSVCVGGDLKSPSTLDMTPNGSYARHGNHTNLGRKDDDEKFYSYHKLSNKFYSYRSP
IRRIKSMSSKYWKQVPKDFGYSRADVGKTLRYKRKSCYGYNAWQREIIYKRRRSPDRS
SVVTSDDGLSSGSVSKLPKKGDTVKLSIKSFRIPELFIEVPETATVGS�KRTVMEAVSVL
LSGGIRVGLMHGKKVRDERKTLSTQGISCDENLDNLGFTLEPSPSKVPLPLCSEDPVP
TDPTSLERSAASPMDSGIPHADDVIDSRNIVDSNLELVPIYQDISVDEPSSDSKELVP
LPELEVKALAIIVPLNQPKRTELAQRRTRRPFSVTEVEALVQAVEELGTGRWRDVKLRAF
EDADHRTYVDLKDQKWLTVHTASISPPQRRGEPVPQELLDRVLRAYGYWSQHKGKHQARG
ASKDPDMNRGGAFESGVSV*

>G892 (21..1004)

TATAACAATTCTTCCAACAATGTTCATTGAGTCAGCCAATAACACGGACCGATAGTGCAC
CCAATGGAGCATTAGGACTTTTGGTCTCTACTGGTGCTACCATTTGTGATCGTATGGTCA
GAATTGCATCCTCTAACCCATCAGAGATCGCCTGTCTCGATGTTTGAGGCAATTTGTCG
TTGAGATTGAAACGAGACAACGGCCTCGGTTTACTTTCAACCATGCTACTCCGCCTTTTG
ATGCTTCTCTGAGGCTCGTCTTCTCGAAGCTCTCTCGCTCATGTTTGAGCCTGCAACCA
TAGGTAGGTTTGGTGACAGACCCATTTCTTAGGGCAAGATCCAGAAACATCTTGGAACCTG
AATCAAGACCCCGACCGCAACATCGAAGACGACACAGCCTTGACAATGTTAACAATGGTG
GTTTACCTCTACCAAGAAGAACATATGTTATTCTCCGGCCCAATAATCCGACTAGTCCAC
TCGGAAACATAAATTGCGCCACCAATCAAGCACCACCACGGCATGTGAACTCACATGATT
ACTTTACTGGAGCATCAAGCTTAGAGCAGCTGATTGAACAACATAACACAAGACGATAGGC
CTGGACCACCACCTGCGTCAGAACCCACCATTAAATCCCTACCATCTGTGAAAATAACAC
CACAACATCTAATAACGACATGTCCCAATGCACAGTGTGCATGGAAGAATTCATTGTTG
GTGGGGACGCAACGGAATTACCATGTAAACATATTTACCATAAAGATTGTATAGTCCCGT
GGCTTAGGCTTAACAATTCTTGCCCTATCTGCCGCGTGACCTGCCACTTGTCAACACCG
TTGCTGAATCTCGAGAAAGGAGCAATCCTATTAGACAAGACATGCCTGAAAGAAGGCGTC
CAAGGTGGATGCAACTCGGTAAACATTTGGCCATTTAGAGCAAGATACCAAAGGGTTAGTC
CAGAAGAAACAGCAAACAGAAATCCTCGAGATAACAGGAGCTAACTCTGAATATTCCATG
GGAAATAAAATCGTGACTATCTATATGTATAGACTCTATGAGACATTGTCTATTTGAAT
GTGCATGTATATCTCAGAAATAAACTCAAGCGAAACATATTTAACGACTAAAAAAA

>G892 Amino Acid Sequence (domain in AA coordinates: 177-270)

MSLSQPITRTDSAPNGAFRTFGLYWYHCDRMVRIASSNPSEIACPRCLRQFVVEIETRO
RPRFTFNHATPPFDASPEARLLEALSLMFEPATIGRFGADPFLRARSNIILEPESRPRPQ
HRRRHS�DNVNNGGGLPLPRRTYVILRPNNPTSPLGNI IAPPNQAPPRHVNSHDYFTGASS
LEQLIEQLTQDDRPGPPPPASEPTINSLPSVKITPQHLTNDMSQCTVCMEEFIVGGDATEL
PCKHIYHKDCIVPWLRLNNSCPICRDLPLVNTVAESRERSNPRIQDMPERRRPRWMQLG
NIWPFRRARYQVSPREETANQNPRDNRS*

>G961 (1..1200)

ATGTCAAAATCTATGAGCATATCAGTGAACGGACAATCTCAAGTGCCTCCTGGGTTTAGG
TTTCATCCGACCGAGGAAGAGCTGTTGCAGTATTATCTCCGGAAGAAAGTTAATAGCATC
GAGATCGATCTTGATGTCAATTCGCGACGTTGATCTCAACAAGCTCGAGCCTTGGGACATT
CAAGAGATGTGTAAATAGGAACAACGCCACAAAACGACTGGTATTTCTTTAGCCACAAG
GACAAAAAATATCCGACGGGAACGAGAATAACAGAGCCACTGCCGCTGGATTTTGGAAA
GCAACTGGCCGCGACAAGATCATATATAGCAATGGCCGTAGAAATTGGGATGAGAAAGACT
CTTGTTTTCTACAAAGGCCGAGCTCCTCACGGCCAAAAATCTGATTGGATCATGCATGAA
TATAGACTCGATGACAACATTATTTCCCCGAGGATGTACCGTTCATGAGGTCGTGAGT
ATTATAGGGGAAGCATCACAGACGAAGGATGGGTGGTGTGTCGTATTTTCAAGAAGAAG
AATCTTCAAAAACCTTAACAGTCCCGTCGGAGGAGCTTCCCTGAGCGGCGGCGGAGAT
ACGCCGAAGACGACATCATCTCAGATCTTCAACGAGGATACTCTCGACCAATTTCTTGAA
CTTATGGGGAGATCTTGTAAGAAGAGCTAAATCTTGACCTTTCATGAAACTCCCAAAC

CTCGAAAGCCCTAACAGTCAGGCAATCAACAACCTGCCACGTAAGCTCTCCCGACACTAAT
CATAATATCCACGTCAGCAACGTGGTCGACACTAGCTTTGTTACTAGCTGGGCGGCTTTA
GACCGCCTCGTGGCCTCGCAGCTTAACGGACCCACATCATATTCAATTACAGCCGTCAT
GAGAGCCACGTGGGCCATGATCATCTCGCTTGCCTTCCGTCCGATCTCCGTACCCACAGC
CTAAACCGGTCCGCTTCGTACCACGCCGGTTTAAACAGGAATATACACCGGAGATGGAG
CTATGGAATACGACGACGTCGTCTCTATCGTCATCGCCTGGCCCATTTTGTACGCTGTCG
AATGTTTTGTGCTTGTGTTGTCTCCTTCGTCTGCAGCTTCAGTTCTGGCCGTTCCAACCA
TGGCAGAGGCAGGTTTCATTTTCGATCTTTCATCGCCTCAGATGCAGATCTCTCTCCATTGA
>G961 Amino Acid Sequence (conserved domain in AA coordinates: 15-140)
MSKSMSISVNGSQVPPGFRFHPTEEELLQYYLRKKVNSIEIDLVDVIRDVDLNKLEPWDI
QEMCKIGTTPQNDWYFFSHKDKKYPTGTRTNRATAAGFWKATGRDKIIYSNGRRIGMRKT
LVFYKGRAPHGQKSDWIMHEYRLDDNIIISPEDVTVHEVVSIIIGEASQDEGWVVCRIFFKK
NLHKTNLNSPVGGASLSGGGDPKTTSSQIFNEDTLDQFLELMGRSCKEELNLDPFMKLPN
LESPNSQAINNCHVSSPDNENIHVSNVVDTSFVTSWAALDRLVASQLNGPYSYSITAVN
ESHVGHDLHALPSVRSPYPSPSLNRSASYHAGLTQEYTPPEMELWNNTTSSLSSSPGPFCHVS
NVLLLVLCLRLQLQFQWPFQWQQRQVHFDLSSPQMQLSLH*
>G1465 (163..1125)
TATCCTTCGCAAGACCCTTCCTCTATATAAGGAAGTTTCATTTGAGAGGACACGC
TGACAAGCTGACTCTAGCTTATCTGGTACCGTCGACCTCATTCTTGCGTTTGATCTTTCT
TTCTCTAGATCCCATATTTTTCTTGATCAATTTAGTTTCATTATGGAGGAAGATGCAGCT
TTTGATCTACTCAAAGCCGAACCTCTAAACGCAGAAGACGATGCAATAATCTCACGTTAT
CTGAAGCGTATGCTCGTCAACGGAGACTCATGGCCTGATCACTTCATCGAAGACGCAGAC
GTGTTCAACAAGAATCCAAATGTGGAGTTCGATGCTGAGAGCCCTAGCTTCGTGATAGTT
AAACCTCGAACAGAGGCTTGTGGTAAAACCGATGGATGTGAAACTGGTTGCTGGAGGATC
ATGGGTGCTGATAAAACCGATAAAATCGACGGAGACTGTGAAGATTCAAGGTTCAAGAAG
ATTCTCAAGTTCTGCCTAAAGAGGAAACCTAGAGGATACAAGAGAAGTTGGGTAATGGAA
GAGTATAGGCTTACCAATAACTTGAAGTGAAGCAAGATCATGTGATTGCAAGATTCCGG
TTTATGTTTGAAGCTGAAATCAGTTTCTTGCTAGCCAAGCATTTCTACACTACATCAGAA
TCACTTCCTCGAAATGAGCTGTTGCCAGCTTACGGATTCTTTTCATCAGATAAGCAATTG
GAGGATGTATCTTATCCGGTGACGATAATGACTTCTGAAGGAAACGATTGGCCTAGCTAC
GTTACCAACAATGTGTATTGTCTGCATCCATTGGAGCTCGTTGATCTTCAAGATCGGATG
TTAATGATTACGGAACCTGCATCTTCGCTAACAAGACTTGTGGTAAAACCGATAGATGC
ATTAATGGTGGTTACTGGAATAATTTGACCCGTGATAGGCTGATCAAGTCAAAGTCCGGG
ATAGTTATTGGTTTCAAGAAGGTGTTAAGTTTCATGAAACGGAGAAAGAAAGATACTTC
TGTGGTGGAGAAGATGTGAAGGTAACCTGGACTCTAGAAGAGTATAGGCTTAGCGTGAAG
CAGAATAAATCTTGTGCGTTATCAAGTTTACTTATGATAACTAAGAATCTTTTCTTGG
ATTTTATGATCATCTTAGTATCGCGACCGCTCTAGACAGGCCTCGTACCGGATCCTCTAG
CTAGAGCTTTCGTTCTGATCATCGGTTTCGACAACG
>G1465 Amino Acid Sequence (conserved domain in AA coordinates: 242-306)
MEEDAADFLLKAEELNAEDDAIISRYLKRNVVNGDSWPDHFIEDADVFNKNPNVEFDAES
PSFVIVKPRTEACGKTDGCETGCWRIMGRDKPIKSTETVKIQGFKILKFLKRPGRYK
RSWVMEYRLTNLNLNWKQDHVICKIRFMFEAEISFLAKHFYTTSESLPRNELLPAYGFL
SSDKQLEDVSYPTIMTSEGNWPSYVTNNVYCLHPLELVDLQDRMFNDYGTICIFANKTC
GKTDRCCINGGYWKILHRDLIKSKSGIVIGFKKVFKEHETEKERYFCGGEDVKVITWLEE
YRLSVKQNKFLCVIKFTYDN*
>G425 (45..1196)
GAAACAGTCTTCTCTTCTCCGATCCCAAAAACGCAGGAAAACAATGTGCTTTAACAGCTCCC
ACCTCCTTCTCCCAAGAAGACCTTCTCTCCGACACTTCACCGATCAATCACAGCAACCTC
CGCCGACGCTCACTTCTCTGAAACACCTTCGCTTGTCAACGCCAGTTTCTCAACCTCCCTA
CCACCCTTACCAGTGGCGATTCCGATCTCGCTCCTCCGACCGCAACGGAGACAATTCCGTT
GCTGATACAAAACCCACGCTGGCTCTCCTTTCATTCCGGAGATGCAAAATACTGGAGAAGTACG
TTCTGAAGTTATCGACGGAGTCAACGCCGATGGTGAACGATACTCGGCGTTGTAGGAGGT
GAAGATTGGCGGAGTGTAGCTATAAGCGCGCATTTTAAGACATCCGATGTACGAGCAGC
TTCTTGCGGCTCACGTGGCTTGCCTTAGGGTTGCGACTCCCGTTGACCAGATTCCGAGGATC
GATGCTCAGCTCAGTCAGTTGCATACCGTCGCCGCGAAATACTCCACTCTTGGTGTGGTTGTT
GACAACAAGGAACCTTGATCATTTTCATGTACATTATGTTGCTTGTATGTTTAAAGAACA
ACTCCAACACCACGTTTGTGTCCATGCAATGGAAGCCATTACCGCTTGTGGGAGATTGAACA

ATCACTGCAATCCCTAACTGGAGTTTCTCCAAGTGAAAGTAATGGTAAGACAATGTCGGATGA
TGAAGATGATAATCAAGTAGAGAGCGAGGTGAACATGTTTGATGGAAGTTGGACGGTTCAG
ATTGCTTGATGGGGTTTGGTCCTCTTGTTCACCGAGAGAGAGATCTTTGATGGAACGTG
TGAAGAAAGAACTGAAGCATGAGCTTAAACAGGGTTTCAAAGAGAAGATTGTGGACATAAG
AGAAGAGATAATGAGGAAGAGAAGGGCGGGAAGTTGCCAGGAGATACGACTTCTGTACT
CAAAGAATGGTGGCGAAGTCACTCGAAATGGCCATACCCAAGTGAAGATAAGGCAAAA
GAACTGGAACAGCAACTCTTCCACGTCATCTACTCTACCAAGAACAACGTAAACGGACC
GGGAAGTCGTAGGTGACATAGCGGCTAACTAGAGGATGGTTCTTTGCCATGTGAATTCTTGG
GAACGTATATGAAAGAAACGAATCCGGTTCTATGCTCGTACAGAGTGTGTTATTTGTATAGT
GGATACCGGTTAGCCTATGAAACCGGATTCGGAGTCCAAATTGTTGTTTGTAAACGACTTAGT
AGTTTTTGAAGTGATCTGTTTCGTTGGTTTGGCTCTTGTAAACGAACGCTTAAGCAAGTGTGGG
TTTTTCTTGTAAAGTGTCAATATGTTTCGTTTGTAAATGAATGTATCAAGCAATATTTATCATAATT
AACTAGCTTGAATGTAAAAA

>G425 Amino Acid Sequence (domain in AA coordinates: TBD)

MSFNSSHLLPPQEDLPLRHFTDQSQQPPQRRHFSETPSLVTASFLNLPPTLTADSDLAPPHR
NGDNSVADTNPRWLSFHFSEMONTGEVRSEVIDGVNADGETILGVVGGEDWRSASYKAAILR
HPMYEQLLAAHVACLRVATPVDQIPRIDAQLSQLHTVAAKYSTLGVVVDNKLDFHMSHYVVL
LCSFKEQLQHHVCVHAMEAITACWEIEQSLQSLTGVSPSESNKGTMSDDEDDNQVESEVNM
FDGSLDGSDDLGMFGPLVPTERERSLMERVKKELKHELKQGFKEKIVDIREEIMRKRRAGKLP
GDTTSLVKEWWRTHSKWPYPTEEDKAKLVQETGLQLKQINNWFINQRKRNWNSNSSTSTLT
KNKRKRTGKS*

>G347 (1..570)

atgaaagtagcagatatgcaggaccagctgggtgtgcatgggtgtaggaatttattgatg
taccctagaggagcatctaattgtgcgttgtgcgttatgtaacactatcaacatgggtcct
cctcctcctccacctcacgacatggcacacattatatgtgggtggttagaacaatgctt
atgtatacgcgtggggctagtagcgttaagatgctcctgtgtcaaaactacgaacctgtg
ccagcgcactccaatcagggttgcccatgctcctccagtcaggttgccgcagatcaattgt
gggcattgtcgagcagacctcatgtatccttacggtgcatcatccgtcaaatgcgctgtt
tgtcaattcgtaactaacgttaatatgagcaatggaaggttacctctcccaactaaccgg
ccaatggaacagcttgccccctctacatcaactcaacaccacctctcagacccaa
accgttggtgtagaaaaccccatgtccgttgatgaaagcggaaggttggtgagcaatggt
ggtgtggagtgacaactgacaaaaagtaa

>G347 Amino Acid Sequence (domain in AA coordinates: 9-39, 50-70, 80-127)

MKVADMQDQLVCHGCRNLLMYPRGASNVRCALCNTINMVP PPPPPHMAHIICGGCRTML
MYTRGASSVRCSCCQTTLNLPVPAHSNQVAHAPSSQVAQINCCHCRTTLMYPYGASSVKCAV
CQFVTNVNMSNGRVPLPTNRPNGTACPPSTSTSTPPSQTVVVENPMSVDESGLVSNV
VVGVTDDKK*

>G1512 (1..732)

ATGGAAGGGAACCTTCTCATCAGGTCTGATGCTCAACGAGCATGACAATGGCTTCATA
GCCAAACAAAAACCTAATCTCACCACGGCTCCAACAGCAGGTCAAGCTAATGAAAGTGGC
TGTTTGGACTGCAACATCTGTTTAGACACAGCCCATGATCCGGTGGTCACTCTCTGCGGG
CACCTTTTCTGCTGGCCTTGCAATTTACAAGTGGTTACATGTTCACTTATCTTCTGTCTCC
GTTGATCAGCACCAGAACAATGCCCTGTTTGTAAATCCAACATTACTATCACCTCTTTG
GTTCTCTCTATGGAAGAGGCATGTCTTCGCCTTCTTCCACGTTTGGCTCCAAGAAACAA
GACGCACTGTCCACTGACATACCCCGCAGACCTGCTCCATCAGCCTTACGCAATCCGATT
ACCTCAGCATCATCTCTGAACCAAGCTTGCAACATCAAACCTCTGCTCCTTCATTTTCA
AATCATCAGTATTCCTCTCGTGGCTTCAACACAACCGAATCAACCGACCTTGCCAATGCT
GTAATGATGAGTTTCTCTACCTGTGATTGGAATGTTTGGAGACCTGGTCTACACCAGG
ATATTGGGACCTTCACAAACACAATAGCTCAGCCTTACCAAAGCCAGAGGATGATGCAG
CGTGAGAAGTCTCTTAATCGGGTATCGATATTCTCTCTTGTGTCATCATCTTTGCCTC
CTTCTCTTCTAG

>G1512 Amino Acid Sequence (domain in AA coordinates: 39-93)

MEGNFFIRSDAQRADHNGFIAKQKPNLTAPTAGQANESGCFDCNICLDTAHPVVTLCG
HLFCWPCTIYKWLHVQLSSVSDQHQNCPVCKSNITITSLVPLYGRGMSSPSTFGSKKQ
DALSTDIPRRPAPSALRNPIITASLNLPSLQHQTLSPSFHNHQSYPGFTTTTESTDLANA
VMMSFLYPVIGMFGDLVYTRIFGTFTNTIAQPYQSQRMMQREKSLNRVSIFLCCIILCL
LLF*

>G2069 (1..1026)

ATGGAAGGAGGAGGAAGAGGACCAATCAAACGATTCTCAGTGAAATAGAACATATGCCT
GAAGCTCCACGTCAACGTATCTCTCATACCGTCGAGCTCGCTCTGAAACCTTCTCTCC
GGCGAATCAATCGACGATCTCCTCTTATTCGATCCTTCGATATCGATTCTCTCTCTA
GACTTCTCTCAACGCTCCACCACCACCACAACAATCACAACAACAACCGCAAGCTTCTCCC
ATGTCCGTTGATTTCGGAAGAAACCTCATCGAACGGTGTGTTCTCTCTAATTCTCTTCT
CCAAAACCCGAAGCTAGATTTCGGTCGCCATGTTTCGTAGCTTCTCGGTTGATTCCGATTTC
TTCGATGATTGGGTGTTACTGAGGAGAAAGTTTATAGCTACAAGTTCAGGAGAGAAGAAG
AAAGGGAATCATCATATAGCAGGAGTAATTCTATGGATGGAGAGATGAGTTCGGCGTCG
TTTAATATCGAATCGATTCTTAGCTTCTGTGAGTGGTAAAGATAGTGGGAAGAAGATATG
GGTATGGGTGGTGATAGACTTGCTGAGCTTGCTTTGCTTGATCCTAAAAGAGCTAAAAGG
ATTTTAGCGAATAGACAATCTGCGGCGAGGTCGAAAGAGAGGAAGATTAGGTATACTGGT
GAGTTAGAGAGGAAGTTTCAGACACTTCAGAATGAAGCTACTACATTGTCTGCTCAAGTC
ACTATGTTACAGAGAGGAACATCAGAGCTGAACACTGAAAATAAACACCTCAAAATGCGG
CTTCAAGCTTTAGAGCAACAAGCTGAACCTTAGGGATGCTTTGAATGAAGCGCTGCGGGAT
GAAGTGAACCGACTTAAGGTGGTAGCTGGAGAAATTCCTCAGGGGAATGGAATTCCTTAC
AACCGTCTCAATTCTCATCTCAGCAATCGGCAATGAATCAGTTTGGGAACAAAACGAAC
CAACAGATGAGTACAAACGGGCAGCCATCGCTCCCAAGCTACATGGATTTACCAAGAGA
GGCTGA

>G2069 Amino Acid Sequence (domain in AA coordinates: TBD)
MEGGGRGNPTILSEIHMPEAPRQRISHHRRARSETFFSGESIDDLFFDPSDIDFSSL
DFLNAPPPPPQSQQQPQASPMVDSEETSSNGVVPNSLPKPEARFGRHVSFSVDSDF
FDDLGVTEBKFIATSSGEKKGNHHHSRNSMDGEMSSASFNIESILASVSGKDSGKKNM
GMGGDRLABLALLDPKRAKRILANRQSAARSKERKIRYTGELERKVQTLQNEATTLSAQV
TMLQRGTSELNTENKHLKMRLLQALEQQALRDALNEALRDELNRLKVVAGEIPQGNNGSY
NRAQFSSQQSAMNQFGNKTNQMSITNGQPSLPSYMDFTKRG*

>G1852 (55..1857)

CATCTGATCTGCTCTCGAAGACGAAAGCTTCGAGTACTGGTTGAAGCTAAAGCTATGGGA
CACGTGAATCTACCTGCATCAAAGCGTGGTAACCCCTCGTCAATGGCGTCTCCTCGACATC
GTAACCGCTGCTTTCTTCGGTATCGTACTTCTCTTCTTCATCCTTTTATTCCTCTCTT
GGTGATTCCATGGCGGCTTCTGGTCGGCAACGCTGCTTCTCTCTACGGCGTCAGATCCG
AGGCAACGGCAGCGATTAGTGACTTTGGTTGAAGCTGGTCAGCATTGCAACCGATCGAG
TATTGTCTGCGGAAGCTGTTGCTCATATGCCCTTGTGAGGATCCGAGAAGGAATAGTCAG
CTTAGTAGAGAGATGAATTTCTATAGGGAGAGACATTGTCTTTCCTGAGGAGACTCCG
CTCTGTTTGATTCTCTCCGCTTCTGGTTATAAAATTCCTGTTCCGTGGCCTGAGAGTCTT
CACAAGATTTGGCATGCAACATGCCATATAACAAAATTGCTGACCGGAAAGGTCATCAA
GGATGGATGAAAGGGAAGGGAATACTTTACTTTCCAGGCGGTGGCAGCATGTTTCTCT
GGCGGAGCTGGCCAATACATTGAAAAGCTTGACAGTATATTCCGCTTAATGGTGGAAC
TTGAGAACTGCTCTTGACATGGGATGCGGGGTAGCTAGTTTGGAGGTAATCTACTATCT
CAAGGCATTCTAGCCCTCTCATTTGCTCCAAGAGATTTCATATAATCTCAAATTCAGTTC
GCTTTGGAAAGAGGAGTGCCTGCTTGTGCTGCTTGGCAGTCTGCTAGACTCCCTTT
CCTGCATACTCTTTGACCTGATGCACTGTTCCCGATGTTTGGATTCCTTTACGGCTTAC
AATGCAACTTACTTTCATCGAAGTAGATAGGTTACTGCGCCCTGGAGGATATCTTGTATC
TCTGGCCACCTGTACAAATGGCCTAAACAAGACAAAGAATGGGCTGATCTTCAGGCGGTG
GCTAGAGCTTTGTGCTATGAGCTAATTGCGGTTGATGGAACACTGTCTGGAAGAAG
CCTGTTGGAGATTATGCTTACCTAGCCAGAATGAGTTTGGGCTGAGTTGTGTGATGAG
TCTGTTCCGCAAGTGATGCATGGTATTTAAATTGAAGAGGTGTGTTACCAGGCCATCA
TCCGTCAAAGGAGATCACGCTTTGGGAACATATCCAAGTGGCCGGAGAGGCTTACTAAA
GTTCTCTTCTAGGGCCATTGTCTGAAAAACGGATTGGATGTGTTGAAGCAGATGCAAGG
CGGTGGGCAAGACGCGTTGCTTATTACAGGGATTCTCTTAACCTGAAGCTGAAATCTCCA
ACTGTCCCAATGTCTAGGACATGAACGCATTCTTCGGAGGCTTTGCAGCAACCCCTTGCA
TCTGATCTCTGTGTTGTTATGAATGTCATTCCAGCTCGGAAGCCATTAATCTTGACGTG
ATTTATGACAGAGTCTCATCGGTGTTTACCATGATTGGTGTGAACATTTTCAACATAT
CCCCGCACGTATGATTTCATCCATGTATCAGGAATTGAATCACTGATAAAACGACAAGAC
TCAAGCAAATCGAGGTGTAGCCTAGTAGATCTAATGGTAGAGATGGACAGAATATTACGT
CCAGAAGGAAAGGTTGTGATCCGAGACTCTCTGAGGTGCTAGATAAAGTCGCACGAATG
GCTCATGCTGTAAGATGGTCTTCTTCCATACACGAGAAAGAACCTGAATCCCATGGAAGA

GAGAAGATTCTTATCGCAACCAATCTCTCTGGAAATTGCCATCAAACCTCCCACTGAAGA
CACAAAAGAAGAAGAAAAGAAGCTCTTCTCAATCTTGTAGGTACTGTCACTTGCTCT
CCAGCCC.

>G1852 Amino Acid Sequence (domain in AA coordinates: 1-601)
MGHVNLPAKRGNNPRQWRLLDIVTAFFGIVLLFFILLFTPLGDSMAASGRQTLTLLSTAS
DPRQRQLVLTVEAGQHLQPIEYCPAEVAHMPCEDPRRNSQLSREMNFYRERHCPLEPE
TPLCLIPPPSGYKIPVPWPESLHKIWHANMPYNKIADRKHGQGMKREGEYFTFPGGGTM
FPGGAGQYIEKLAQYIPLNGGTLRLTALDMGCGVASFGGTLTLLSQGILALSFAPRDSHKSQI
QFALERGVPAFVAMLGTRRLPFPAYSFDLMHCSRCLIPFTAYNATYFIEVDRLLRPGGYL
VISGPPVQWPKQDKEWADLQAVARALCYELIAVDGNTVIWKKPVGDSCLPSQNEFGLLELC
DESVPSPDAWYFKLRKCVTRPSSVKGEHALGTISKWPERLTKVPSRAIVMKNGLDVFEAD
ARRWARRVAYYRDSNLNLKLSPTVRNVMDNAFFGGFAATLASDPVWVMNVI PARKPLTL
DVIYDRGLIGVYHWDCEPFSTYPRTYDFIHVSGIESLIKQDSSKSRCSLVDLMVEMDRI
LRPEGKVVIRDSPEVLDKVARMAHAVRWSSSIHEKEPESHGREKILIIATKSLWKLPNSH
*

>G1793 (59..1783)
AGTGATTTATTGATTAACCCAAACACAAAATAAACAGATTTGACTCAAAAAGAAGAAAAT
GAATTTCTAACAACTGGCTTGGCTTTCTCTTTTACCGAACAACTCTTCTTTGCCTCCTCA
TGAATACAACCTTGGCTTGGTCAGCGACCATATGGACAACCC'TTTTCAAACACAAGAGTG
GAATATGATCAATCCACACGGTGGAGGAGAGATGAAGGAGGAGAGGTTCCAAAAGTGGC
CGATTTTCTCGGTGTGAGCAAACCGGACGAAAACCAATCCAACCACCTAGTAGCTTACAA
CGACTCAGACTACTACTTCCATACCAATAGCTTGATGCCTAGCGTCCAATCAAACGATGT
CGTTGTAGCAGCTTGTGACTCCAATACTCCTAACAAACAGTAGCTATCATGAGCTTCAAGA
GAGTGCTCACAATCTACAGTCACTTACTTTGTCCATGGGGACCACCGCTGGTAATAATGT
TGTGATCAAAGCTTCCAGATCCGAGACACCGGGGATAACGCTAGCGGTGGAGCACTAGC
CGTTGTTGAGACGGCCACGCCAAGACGTGCATTGGACACTTTCGGACAACGAACCTCGAT
CTATCGTGGTGTCAAGACATCGATGGACTGGTCGATATGAGGCTCATCTATGGGATAA
TAGTTGTAGAAGGGAAGGCCAGTCTAGGAAAGGAAGACAAGTTTACTTGGGTGGATATGA
CAAAGAAGATAAAGCAGCAAGATCATATGATCTAGCTGCACTTAAGTACTGGGGTCTCTTC
AACTACTACTAATTTCCCCATTACAACTACGAGAAAAGAAGTAGAGGAAATGAAGCATAT
TACGAGACAAGAGCTTCGTGGCTGCCATTAGAAGGAAAAGTAGTGGATTTTCGAGAGGCGC
TTCGATGTATCGAGGAGTTTACAAGGCATCACCAACATGGAAGATGGCAAGCAAGGATCGG
CCGAGTCGCCGGAACAAGACCTCTACTTGGGAAC'TTTTAGCACTGAGGAAGAAGCAGC
AGAAGCTTACGATATAGCTGCAATAAAGTTTAGAGGACTTAATGCAGTGACCAACTTCGA
GATCAACCGGTACGACGTGAAAGCCATTCTAGAGAGTAGCACTCTTCCCATCGGAGGAGG
CGCAGCTAAACGGCTCAAAGAAGCTCAAGCTCTTGAGTCTTCAAGGAAACGCGAGGCCGA
GATGATAGCCCTTGGTTCAAGTTTCCAGTACGGTGGTGGCTCGAGCACAGGCTCTGGCTC
CACCTCATCAAGACTTCAGCTTCAACCTTACCCTCTAAGCATTCAACAACCATTAGAGCC
TTTCTATCTCTTTCAGAACAAATGACATCTCTCATTACAACAACAACAAATGCTCACGATTC
CTCCTCTTTTAATCACCATAGCTATATCCAGACACAACCTTCATCTCCACCAACAGACCAA
CAATTACTTGCAGCAACAGTCGAGCCAGAACTCTCAGCAGCTCTACAATGCGTATCTTCA
TAGCAATCCGGCTCTGCTTCACTGGACTTGTCTCTACCTCTATCGTTGACAACAATAATAA
CAATGGAGGCTCTAGTGGGAGCTACAACACTGCAGCATTTCTTGGGAACACGGTATTGG
TATTGGGTCCAGCTCGACTGTGGATCGACCGAGGAGTTTCCAACCGTTAAAACAGATTA
CGATATGCCCTTCCAGTGATGGAACCGGAGGGTATAGTGGTGGACCAGTGAGTCTGTTCA
GGGGTCAAACCTGGTGGTGT'TTTCATATGTGGAATGAGTAAACAAGGATCTCTTTCTT
GCGGCACAAGGAATGGGT

>G1793 Amino Acid Sequence (conserved domain in AA coordinates:179-255, 281-349)
MNSNNWLGFPLSPNNSSLPPEYNLGLVSDHMDNPFQTQEWNMINPHGGGGDEGGEVPKV
ADFLGVSKPDENQSNHLVAYNDSDYFHTNSLMPSVQSNVVAACDSNTPNNSSYHELQ
ESAHNQLSLTSMGTTAGNNVVDKASPSETTGDNASGGALAVVETATPRRALDTFGQRTS
IYRGVTRHRWTGRYEAHLWDNSCRREGQSRKGRQVYLGGYDKEDKAARSYDLAALKYWGP
STTINFPIITNYEKEVEEMKHMTRQEFVAAIRRKSSGFSRGASMYRGVTRHHQHGRWQARI
GRVAGNKDLYLGTFFSTEEEAABAYDIAAIKFRGLNAVITNFEINRYDVKAILESSTLPIGG
GAAKRLKEAQALESSRKREAEMIALGSSFYGGGSSSTGSGSTSSRLQLQPYPLSIQQPLE
PFLSLQNNDISHYNNNNNAHDSSSFNHHSYIQTQLHLHQQTNNYLQQQSSQNSQQLYNAYL
HSPNALLHGLVSTSIVDNNNNNGGSSGSYNTAAFLGNHIGIGSSSTVGSSTEEFPTVKTD

YDMPSSDGTGGYSGWTSESQVGSNPGGVFTMWNE*
>G761 (521..1549)
GGGGCCGACCGCCGCCCGGGCAGGTCTAGGTTCAAAAGGACTCACAAGAGAGAGATAGT
ATGATTGATAGGGAAAGAGAGAGAGATGAAAGAAAGTAAATATATAATAGATTATTAGG
ACACGAGTGTCTATCTTTTGTATTTGTGTCTTGTGTCTCTCTTTCTCTCTTCTCGAA
TGATCATCTTTATATAACCCCTACTCTCTTTCTCTTTTCCCATTTCTTTCATATCATTCTCC
CTTCTCTCTCTCGGATCTGATCTCTCTTTCCAGTAACCTATTCCCGAGGAGCACTGTCAA
ATCTTGTCCACTCTTTGATCTTATCTCGATCTCTTTCTCTTTCTAGTCTTGTGTAGTCTT
CAAACCTTGTGATGTTATCTATATAGTAATCACGAGAGAGAATCATACAATAGCTGAAACA
TAAAGCTTTCTTAGAAGCTTTAAAAAGGTCTCATCTGGATTATCCTGTTAATTTCTAGA
GTTTCTTCAGGCAGATTATTAACCGATCAAGAAGACAAACATGAATTCATTTTCCACGT
CCCTCCGGGTTTATAGATTTACCCGACAGATGAAGAACCTGTAGACTACTACCTGAGGAA
AAAAGTCGCATCGAAGAGAATAGAAATTGATTTCAATAAGGACATTGATCTTTACAAGAT
TGAGCCATGGGACCTTCAAGAGTTGTGCAAAATTGGGCATGAAGAGCAGAGTGATTGGTA
CTTCTTTAGCCATAAAGACAAGAAGTATCCACAGGACTCGAACCAATAGAGCAACAAA
AGCAGGGTTTGGAAAGCCACCGGAAGAGATAAGGCTATCTATTTGAGGCATAGTCTAAT
TGGCATGAGGAAAACACTTGTGTTTTACAAGGGAAGAGCCCCAAATGGACAAAAGTCTGA
TTGGATCATGCAGAAATACCGCTTAGAAACCGATGAAAACGGAACCTCCTCAGGAAGAAGG
ATGGGTTGTGTGTAGGGTTTCAAGAAGAGATTGGCTGCAGTTAGACGAATGGGAGATTA
CGACTCATCCCTTACATTGGTACGATGATCAACTTTCTTTTATGGCTCCGAGCTCGA
GACAAACGGTCAACGACGGATTCTCCCAATCATCATCAGCAGCAGCAGCAGCAGCAGCA
ACAACATATGCCATATGGCCTCAATGCATCTGCTTACGCTCTCAACAACCTAAGTTGCA
ATGCAAGCAAGAGCTAGAACTACACTACAACCACTGCAATCAAAATATCGCGCATGAGGA
ACAATTGAATCAAGGAAATCAGAACTTCAGCTCTCTATACATGAACAGCGGCAACGAGCA
AGTGATGGACCAAGTCACAGACTGGAGAGTTCTCGATAAAATTTGTTGCTTCTCAGCTAAG
CAACGAGGAGGCTGCCACAGCTTCTGCATCTATACAGAATAATGCCAAGGACACAAGCAA
TGCTGAGTACCAAGTTGATGAAGAAAAAGATCCGAAAGGGCTTCAGACATGGGAGAAGA
ATATACTGCTTCTACTTCTTCGAGTTGTCAGATTGATCTATGGAAGTGAGCTGAAAGAGA
AGACATATAAATGCATATATACATATATATATACGTACACAGCACTAATCAAGTG
TAGATGATGATGATGGTACAGATTATATATTTGCTTTGATTGATTCTTACTACATTATTGA
ACTTATGTCATATGCATATATACATTGCGTATCTATGCATATTTATACTTGTACTCAATA
TGATTAACCATATATAAACTCTAATCTAAATGTAACCTCAATATTTTAAATAGACAAT
TGTCTCTTCTTATTAGAAAAAAA
>G761 Amino Acid Sequence (domain in AA coordinates: 10-156)
MNSFSHVPPGFRFHPDDEELVDYYLRKKVASKRIEIDFIKDIDLYKIEPWDLQELCKIGH
EEQSDWYFFSHKDKKYPGTGRTNRATKAGFWKATGRDKAIYLRHSLIGMRKTLVFKGRA
PNGQKSDWIMHEYRLTDENGTPQEEGWVVCRVFKRLAAVRRMGDYDSSPSHWYDDQLS
FMASELETNGQRRILPNHHQQQHEHQHMPYGLNASAYALNPNLQCKQLELHYNHLQ
SNIAHEQLNQGNQNFSSLYMNSGNEQVMDQVTDWRVLDKFVASQLSNEEAATASASIQN
NAKDTNSAEYQVDEEKDPKRASDMGEEYTASTSSSCQIDLWK*
>G1056 (10..798)
GCTACATATATGGGTTCTATTAGAGGAAACATTGAAGAGCCTATATCTCAGTCATTAACG
AGGCAGAACTCTCTCTATAGCTTAAAGCTCCATGAGGTTCAAACCCACTTAGGAAGTTCT
GGAAACCACTAGGAAGCATGAACCTTGATGAGCTTCTCAAGACTGTCTTGCCACCAGCT
GAGGAAGGGCTTGTTCGTGAGGGAAGCTTGACGTTACCTCGAGATCTCAGTAAAAAGACA
GTTGATGAGGTCTGAGAGATATCCAACAGGACAAGAATGGAAACGGTACTAGTACTACT
ACTACTCATAAGCAGCCTACACTCGGTGAAATAACACTTGAGGATTTGTTGTTGAGAGCT
GGTGTAGTGACTGAGACAGTAGTCCCTCAAGAAAATGTTGTTAACATAGCTTCAAATGGG
CAATGGGTTGAGTATCATCATCAGCCTCAACAACAAGGGTTTATGACATATCCGGTT
TGCGAGATGCAAGATATGGTGTATGATGGGTGGATTATCGGATACACCACAAGCGCTGGG
AGGAAAAGAGTAGCTGGAGAGATTGTGGAGAAGACTGTTGAGAGGAGACAGAAGAGGATG
ATCAAGAACAGAGAATCTGCAGCAGCTTACGAGCTAGGAAACAGGCTTATACATGAA
TTAGAGATCAAGGTTTCAAGGTTAGAAGAAGAAAACGAAAACCTTCGGAGGCTAAAGGAG
GTGGAGAAGATCCTACCAAGTGAACCACCACCATCTAAGTGAAGCTCCGGCGAACA
AACTCTGCTTCTCTGATCTTAAAGACTCTTCTTTCTTCTTCTTCTTGTGTTGGTTT
ATATCAGACCGCTTGTCTTGTATATTGTGTAGACTTTATTGACTTTGAACAGCATGT
CTTTATAACATTTCTTGAGTGT

>G1056 Amino Acid Sequence (domain in AA coordinates 183-246)
MGSIRGNIEEPISQSLTRQNSLYSLKLHEVQTHLGSSGKPLGSMNLDELLKTVLPPAEEG
LVRQGSLLTPRDLSSKKTVDVWRDIQQDKNGNGTSTTTTHKQPTLGEITLEDLLLRAVGV
TETVVPQENVNIIASNGQWVEYHHQPQQQGFMTYPVCEMQDMVMMGGLSDTFPQAPGRKR
VAGEIVEKTVERRQKRMIKNRESAARSARKQAYTHELEIKVSRLEEENEKLRLKEVEK
ILPSEPPDPKWKLRRTNSASL*

>G1447 (82..1086)

AAAAACCCCTAACCCCTAATCTCTCAAGACAACCTCAAAGGTCTCTCCTTTTATAGGTTTAT
TATCACTTCCGTATAATCGCCATGTCTTCTTACCATGGAAAAAACAAAATCGAGTCGA
ATCTTAAGATTCATTTCTGAGTTTCAACAATCACCGTTCGTTGAAACTGGCTTTCCAAC
TCTCTGATCGATCTCTTCTCAAGAATCGCGATCGTCTAAAAAATCTCCATCTAAACGC
TTCCAACGAATCGAACGCCAGATTCTGAACCGCTCCAAACGCTTCTTCTGAGTAATCAA
GATACGATTTTGGAAAGCCCTCGAGGATTAAAAACGTTCTGAAGTAAGTTCGAGAAAGTT
AATTGCGTTAAAGGTAAATCAGCGCGCTTGAAGAAGAACGCGATTAAAAATAGCGTTTTC
GGCGGTAGCCGTGAGGTCGTTTGGATGGCGTTTAAAGTTTGGATAGTAGCGTTGCTCGCC
TTGAGCACGAAGAAGAAGCTCACTTTAGGAATCACTCTCTGCGCTTCGCTCTTCTCTTA
ACAGAGCTCGTGGCGCGCGTGTTCACGCGCTCTAATAACACCGACAAAGACAAAAAC
GCGATTGCCCGCGAGAAAAATCGAACTTTTGGATGAACTCGAGTTCCCAAAGCGATTCCA
TGTCCTGAGGAAACAGAGCATGTAGTATCTGAAACAGAGGTTTGAAGTTGAAAGGTTTA
ACGATACGTGATCTGTTGTCAAAGGACGAGAAATCAACAAGTAAAGTTGGAGACTAAAA
TCGAAGATTGTGAAGAAGTTGAGGAGTTACAATAAGAAGGATAAGAAGACGATGAAGATC
AAAGAAGAGCTTTTGGATTGAAGTCTCGAGTTTGGTTTTAGAAGATAAACCAAAGAAAT
GAGTCTGAGAGACGAAGAAGAAACGTTGAATCCTCCAGTGGTTGGATCAAACCTGAAT
GGGATTGTTCTGATCGTGATGTGCTAACCGGTTTGTATGTGGGAAGGCTTTAGCTATT
GTTCTGACACTATCATGTTTGGTTCTTAGATTAGGAGCAGTCAAAAAAGTTAATCTTTC
ATATAATTTTTTTTGTATTTTTTAACATGCTTGCATGTGAACTGTAAATTTTTCTCATT
CATATGAAGGAGATTGGATTGAATGTTGAATACTAA

>G1447 Amino Acid Sequence (domain in AA coordinates: 3-54, 124-156)

MSSLPWKKPKSSRIIRFISEFQQSPFVETGFPSTLIDLFFKNRDLKKSPSKRFQRIERQ
IRTAPNASSLSNQDTIFEKPSRIKTVRSKVEKVCVKGSAAALKNAIKNSVFGGSGEVV
LMAFKVLIVALLALSTKKKLTGLITLSAFALLLTCLVAAARVFRSNNTDKDKNAIAREKI
ETFDETRVPKAIKPCPEETEHVVSETEVSKLKLGLTIRDLSSKDEKSTSKSWRLSKIVKKL
RSYNKKDKTKMKIKEESLIEVSSVLEDPKKIESERDEEETLNPVVGSNLNGIVLIVI
VLTGLLCCGKVLAIVLTLSCLVLRGLGAVKKVNLCI*

>G323 (77..826)

CTGCTCATATCAGCCATTGACACAGTTGCTTTGGGTTTCCCTCAAACGGCGCCGATTGTC
TGGATTTTGACCCTGATGGCCTTAGATCAATCTTTGAAGATGCTGCTTTACTTGGAGA
ACTCTATGGAGAAGGTGCATTTTGTTCAGAGCAAGAAACCTGAACCCATTACAGTCTC
GGTTCCTTCTGATGATACTGATGATTGCAATTTTGAAGTCAATATTTGCTTAGACTCGGT
GCAAGAACTGTTGTGACTCTCTGTGTCACCTCTTTTGGTGGCCTTGATTACAAATG
GCTTGATGTACAGAGCTTCTCAACAAGTGATGAATACCAAAGACATAGACAGTGTCTGT
TTGTAAATCTAAAGTTTCTCATTCTACTTTGGTTTCTTTGTATGGTAGAGGCCGTTGTAC
TACTCAGGAGGAAGGTAAAAACAGTGTGCCTAAAAGACCCGTAGGACCGGTTTATCGGCT
TGAAATGCCGAATTCACCTTATGCAAGTACTGATCTGCGGTTATCACAACGGGTTTATC
CAATAGCCACAGGAAGGTTACTACCTGTCTCAGGGGTGATGAGCTCGAACAGTTTATC
ATACTCTGCTGTTTTGGATCCGGTGATGGTGATGGTTGGAGAAATGGTAGCTACGAGGT
GTTTGGAAACACGAGTGATGGATAGATTGCGTATCCGACACTTACAATCTCGCAGGGAC
TAGCGGGCCGAGGATGAGAAGGCGGATAATGCAGGCAGATAAATCGCTGGGAAGAATCTT
CTTCTTCTTTATGTGTTGTGTTGTTCTGTGTCTTCTTGTGTTTATAGGTTTTTCATAGCTAG
CTTGGTTCTGCTACTGTTCAAGTTTCTTCAGG

>G323 Amino Acid Sequence (conserved domain in AA coordinates: 48-96)

MALDQSFEDAALLGELYGEGAFCKSKKPEPITVSVPSDDTDDSNFDCNICLDSVQEPVV
TLCGHLFCWPCIIHKWLDVQSFSSTSEYQRRHQCPVCKSKVSHSTLVPLYGRGRCTTQEEG
KNSVPKRPVGPVYRLMPNSPYASTDLRLSQRVHFNSPQEGYYPVSGVMSSNSLSYSAVL
DPVMVMVGEMVATRLFGTRVMDRFAYPDTYNLAGTSGPRMRRIMQADKSLGRIFFFMC
CVVLCLLLF*

>G176 (41..1606)

AGAAGAAGAAGAAGAGTACCTCATACGTAAACCATTGATGGGCTCTTTTGATCGCCA
AAGAGCTGTTCCGAAATTCAAAACAGCAACACCGTCACCGCTCCCTCTTTCTCCTTCGCC
TTACTTCACTATGCCTCCTGGCCTTACTCCCGCCGACTTCTCGACTCTCCTCTTCTCTT
CACTTCCCTCCAAACATTTTGCCGTCTCCTACGACAGGCACATTCCAGCGCAATCTCTGAA
CTATAACAATAACGGTTTGTCTATTGACAAAAATGAAATCAAATATGAAGACACAACCTCC
TCCCTTGTTCCTACCATCTATGGTAACTCAGCCTTTACCTCAACTGGATTATTCAAATC
CGAAATCATGTCGAGTAACAAAACCTCTGATGACGGCTACAATTGGCGCAAATACGGGCA
GAAGCAAGTCAAAGGAAGCGAAAACCCGAGGAGTTACTTCAAATGCACGTATCCAAATTG
TCTCACAAGAAGAAAGTAGAGACGTCTCTTGTAAGGGTCAGATGATTGAGATTGTCTA
TAAAGGAAGCCACAATCATCCCAAGCCCCAATCCACGAAGCGATCATCTCCACCGCTAT
AGCAGCACATCAGAACAGCAGTAATGGAGACGGTAAAGACATTGGTGAAGATGAAACAGA
GGCCAAGAGATGGAAGAAGAGAAGATAATGTGAAGGAGCCAAAGAGTGGTGGTTCAGACAAC
AAGTGATATAGACATTCTTGACGATGGCTACAGATGGAGAAAGTATGGTCAGAAAGTCGT
CAAGGGTAATCCAAATCCAAGGAGCTATTACAAGTGCACATTTACAGGATGTTTTGTAAG
GAAACACGTTGAAAGAGCATTTCAGATCCCAAGTCAGTGATCACAACCTACGAAGGAAA
ACACAAACACCAAATCCCGACCCCAAGAAGAGGTCCAGTTTAAAGATCTGCTGCAATGGC
TTCTCCTCTTCTCCAACTTCGACTACTCTGATCAACTTCCCGGCGGCGATCCACAGTT
GCTGAGCTCTCTACGCGTCCCTCTGTCCCGGCTTCTAGCCACCGTCCGTCACGCTTCTGC
AGATGCCAGACCCTGGGCAGAGCTCGTTGACCGGTGAGCGTTTCCCGGCCACCATCGCT
CTCGGAGGCAACGTACAGTAAGGAAGAATTTTCTATTTCGAGCCAATTACATAAC
CTTAGTGGAATCTTACTCGCCGCGTCTCTGCTCAGCACCCCTTTCGCTCTCTTCTCCT
CGCATCGCTGGCCGCTTCTTGGCTTTTCTCTACTTTTTCCGTCCGGCGGATCAGCCGTT
GGTCATTGGAGGACGCAGTTCTCCGATCTTGAGACGCTAGGGATACTCTGCCTGTCCAC
TGTGGTGGTGATGTTTATGACCAGCGTTGGATCGCTCTTGATGTCCACTCTAGCAGTTGG
GATCATGGGCGTGGCCATCCACGGAGCGTTTCGTGCTCCCGAAGACCTGTTTCTTGAAGA
ACAAGAAGCCATTGGATCTGGACTTTTTCGATTTCTTCAACAACAATGCCTCTAATGCAGC
TGCCGCTGCCATAGCCACCTCAGCAATGTACGCGTTCGAGTCTGAGATTGTTGAAGAGA
CTACATTCCTACACCGCATTTCCAAAGTGTGATATTTATTATATTGAATTGTT
>G176 Amino Acid Sequence (domain in AA coordinates: 117-173,234-290)
MGSFDRQRAVPKFKTATPSPLPLSPSPYFTMPPLTPADFLDSPLLFSSNILPSPTTGT
FPAQSLNLYNNGLLIDKNEIKYEDTTPPLFLPSMVTQPLQLDLFKSEIMSSNKTSDGY
NWRKYGQKQVKGSENPRSYFKCTYPNCLTKKKVETSLVKQMIEIVYKGSNHPKPKQSTK
RSSSTAIAAHQNSSNGDGKDIGEDETEAKRWKREENVKEPRVVVQTSTDIDILDDGYRWR
KYGQKVVKGNPNPRSYKCTFTGCFVRKHVERAFQDPKSVITTYEGKHKHQIPTPRRGPV
LRSAMASPLLPSTTTPDQLPGGDPQLLSLRVLLSRVLATVRHASADARPWAELVDRSA
FSRPPSLSEATSRVRKNFSYFRANYITLVAILLAASLLTHFPALFLLASLAASWLFYFF
RPADQLVIGGRTFSDLETGLILCLSTVVMFMTSVGSLLMSTLAVGIMGVAIHGAFRAP
EDLFLEEQEAIGSGLFAFFNNNASNAAAAAIATSAMSRVRV*
>G174 (194..1585)
CCCAATTTGAGATTGTTTCGATTTCGATCTACGAGATTCTTACAAGAACATAAGCAGCTTC
GGTTTTTTGGGATTATCTTATTTGGTTCGGATGATGATCTTCTCGATGTCTGTGCTAGGCT
TTGGGAATTAGATATATTTGGGGTTAAGCTCGAGTCTCTCCGGTTTTGAGTTTACTTGAG
TTTGTTAGTATTTATGGCTGAGGTGGGAAAAGTTCTGGCTAGTGATATGGAGTTAGACCA
TTCAAATGAGACTAAAGCAGTGGATGATGTTGTTGCCACTACTGATAAAGCGGAGGTCAT
ACAGTGGCTGTAACTAGAAGTGAACCGTTGTTGAAAGTTTGAATCTACTGACTGTAA
GGAGCTTGAAAAAATTTGTTCCACATACGGTAGCTTCGCAGTCGGAAGTAGATGTTGCTTC
CCCGGTATCCGAGAAAGCACCGAAGGTTTCTGAAAGTAGCGGTGCATTATCTTTGCAGTC
TGGTTCGGAAGGGAATAGTCTTTTATTCTGTGAGAAGGTTATGGAAGACGGATACAACCTG
GCGGAAATATGGACAGAACTTGTGAAAGGAAATGAGTTTGTAAGGAGCTATTACAGGTG
CACTCACCCCTAACTGCAAGCGGAAAAAACAGTTGGAACGGTCTGCGGGTGGACAAGTCGT
GGATACCGTTTACTTTGGGGAAACATGATCACCCAAAGCCTCTTGCTGGTGCTGTTCTCTAT
CAATCAGGATAAGCGAAGTGATGTCTTCACAGCTGTTAGTAAAGAGAAAAACATCTGGATC
CAGTGTTCAGACACTTCTCTCAAACCGAACCACCAAGATCCATGGAGGATTACATGTTTC
AGTTATTCCACCAGCTGATGATGTGAAACTGATATTTACAATCAAGTAGGATAACGGG
GGACAACACTCACAAGGATTATAATAGTCTTACCGCAAAGCGAAGGAAGAAAGGAGGGAA
CATTGAGCTGAGTCCAGTGGAGAGGTCAACCAATGATTACGCATTGTGGTTCACACTCA
GACTCTGTTTGATATTGTGAATGATGGTACCGATGGCGTAAATATGGTCAGAAATCAGT

AAAAGGCAGCCCATATCCAAGGAGCTACTATAGATGTTCAAGCCCTGGATGCCCCGTCAA
GAAACACGTAGAGAGGTCTATCTCATGACACAAAGTTGCTTATAACAACCTACGAGGGAAA
ACACGACCACGATATGCCTCCAGGAAGAGTTGTTACTCATAATAACATGCTGGACTCGGA
AGTTGATGATAAAGAAGGAGATGCCAACAAGACTCCACAGAGCTCAACTCTTCAATCCAT
TACAAAAGACCAGCATGTGCAAGATCACTTAAGAAAAGAAAACGAAGACTAATGGCTTTGA
GAAAAGTCTTGATCAAGGTCCAGTTTGGATGAGAAGCTGAAGGAGGAAATAAAAGAGAG
ATCAGATGCAAAACAAAGATCACGCAGCCAATCACGCCAAGCCGGAAGCAAAGTCAGATGA
TAAACCCTGTTTGTCAAGAGAAGGCAGTAGGAACCCTGGAGAGCGAGGAACAAAACC
CAAGACAGAGCCTGCCCAAAGCTAAGCATTCAAGTGTGTACCGAGTGGTAATTTATATGG
CTGTTTAAACATAGATTAGTACAGGCGATATGGTTATAGACTGTACAGTTGTTGTTTCAGG
CGGGACCAGATTAGATTAGTGTAAATGGAATAGTATGCTTAAATACCTTTATGTAACC
ACTTCCATTTGGTTCAAATAAGAGTTACAGGAAGAGAAGGTAACACAACAAGAGCCCTTC
TTTGTGATGGAGCCTGTGTAATAGTTGTAGCATGGGGATGTATATGATTGATTCAACC
TTATTAATGGTTATGAGACAAAACCTATC

>G174 Amino Acid Sequence (domain in AA coordinates: TBD)
MAEVGKVLASDMELDHSNETKAVDDVVATTDKAEVIPVAVTRTETVVESLESTDCKELEK
LVPHTVASQSEVDVASPVSEKAPKVSSESSGALSLSQSGSEGNPFIREKVMEDGYNWRKYG
QKLVKGNFVRSYYRCTHPNCKAKKQLERSAGGQVVDTVYFGEHDHPKPLAGAVPINQDK
RSDVFTAVSKEKTSGSSVQTLRQTEPPKIHGGLHVSVIPADDDVKTDISQSSRITGDNTH
KDYNSTAKRRKKGGNIELSPVERSTNDSRIVVHTQTLFDIVNDGYRWRKYGQKSVKGS
YPRSYRCSSPGCPVKKHVERSSHDTKLLITTYEGKHDHDMPPGRVTVTHNMLDSEVDDK
EGDANKTPQSSTLQSIITKQHVHDLRKKTKTNGFEKSLDQGPVLDEKLKEIKERSDAN
KDHAAHNAKPEAKSDDKTTCVQEKAVGTTLESEEQKPKTEPAQS*

>G715 (1..705)

ATGGATACCAACAACCAGCAACCACCTCCCTCCGCCGCCGAATCCCTCCTCCACCACCT
GGAACCACCATCTCCGCCGAGGAGGAGGAGCTTCTTACCACCACCTTCTCCAACAACAA
CAACAACAGCTCCAATATTCTGGACCTACCAACGCCAAGAGATCGAACAAGTTAACGAT
TTCAAAACCATCTAGCTTCCACTAGCTAGGATAAAAAAGATCATGAAGCCGATGAAGAT
GTTGCGTATGATCTCCGCGAGAAGCACCGATTCTCTTCGCGAAAGCTTGTGAGCTTTTCATT
CTCGAGCTCACGATCAGATCTTGGCTTCACGCTGAGGAGAATAAACGTCGTACGCTTCAG
AAAAACGATATCGCTGCTGCGATTACTAGGACTGATATCTTCGATTCTCTTGTGATATT
GTTCTTAGAGATGAGATTAGGACGAAGCCGAGTCTCGGTGGTGAATGGTGGTGGCT
CCTACCGCGAGCGCGCTGCTTACTATTATCCGCCGATGGGACAACCAGCTGGTCTCTGGA
GGGATGATGATTGGGAGACCAGCTATGGATCCGAATGGTGTATGTCCAGCCTCCGTCT
CAGGCGTGGCAGAGTGTGTCGACACTTCGACGGGACGGGAGATGATGTCTTATGGT
AGTGGTGGGAAGTTCCGGTCAAGGGAATCTCGACGGCCAAGGTA

>G715 Amino Acid Sequence (domain in AA coordinates: 60-132)
MDTNQPPPPSAAGIPPPPGTTTISAAGGGASYHLLQQQQQLQLFWTYQRQEIEQVND
FKNHQLPLARIKKIMKADEVDVRMISAEAPILFAKACELFLELTIRSWLHAEENKRRTLQ
KNDIAAAITRTDIFDFLVDIVPRDEIKDEAAVLGGGMVAPTASGVPIYYPPMPGQAPGP
GMMIGRPAMDPNGVYVQPPSQAWQSVWQTSTGTGDDVSYGSGSSGQGNLDGQG*

>G588 (196..1599)

ATCTGAAGTGAACCAAGCTCAGGTTTTGTCTTCTCTTGTATCATTCCTTTCTCAGCAATA
TAAATTAGAGTTATATCCTTTTATAAAGGATTTTGCTTTTTCACCAACAAACCCTAAATTC
GGTGTCTCAGCAAGAATCACGTGATTCTCGTTCTCTCTCTCACGAAACCCATCATCTTC
TATCTCATTGAGAAAATGGGTCAAAAGTTTTGGGAGAATCAAGAAGATCGAGCGATGGTT
GAATCCACCATAGGCTCTGAAGCTTGCAGCTTTTTCATCTCAACAGCTTCAGCTTCCAAC
ACTGCCCTTGTCCAAGCTTGTCTCACCAACAAGTGATTCCAATCTCCAACAAGGGTTACGT
CAGTTTGTGAAGGATCTGATTGGGATTATGCTCTTTTCTGGCTAGCGTCCAACGTTAAT
AGCTCTGATGTTGTCTCTTGATCTGGGGAGATGGTCATTGCCGTGTCAAAAAGGGTGCT
TCAGGTGAGGATTACTCTCAGCAAGATGAGATCAAAAGACGTGTGCTTCGCAAGCTTCAC
TTGTCTGTTCTGTTTCAGATGAAGATCATCGTTTGGTGAAATCAGGAGCTCTTACTGAT
CTCGACATGTTTTATCTGGCTTCTTTGTACTTTTCTTTAGGTGTGATACCAATAAGTAC
GGTCTGCTGGAACCTATGTGTCTGGGAAGCCTCTTGGGCTGCAGATTTGCCTAGCTGC
TTGAGTTATTATAGGGTTAGGTCTTTCTTAGCTAGGTGAGTCTGTTTTTCAGACTGTGTG
TCTGTACCAGTGAATCTGGAGTTGTGGAGCTTGGTTCTTTAAGACATATTCAGAAGAT
AAGAGTGTGATTGAGATGGTGAAATCAGTGTGGTGGGCTGACTTTGTTTCAGGCTAAA

GAAGCTCCTAAAATCTTTGGTTCGACAGCTGAGTCTTGGTGGAGCAAAACCTCGGTCTATG
AGTATTAATTTCTCCCCGAAGACCGAGGATGACACGGGTTTCTCATTGGAATCGTATGAG
GTGCAAGCGATCGGAGGCTCTAATCAAGTGTATGGTTATGAGCAAGGGAAAGATGAGACA
TTGTATCTAACTGACGAGCAAAAGCCGAGGAAGAGAGGGAGAAAACAGCAAATGGAAGA
GAAGAGGCTCTAAACCATGTGTGAAGCGGAACGGCAGAGGAGGAGAAGCTGAACCAGAGA
TTCTACGCTTTGAGAGCGGTGGTGCTAACATCTCCAAGATGGACAAGGCTTCGCTCCTT
GCAGACGCAATCACTTACATCACGGATATGCAGAAGAAAATCAGGGTGTATGAAACAGAG
AAGCAGATAATGAAGAGGAGGAGAGTAATCAGATAACTCCAGCAGAGGTTGATTATCAA
CAGAGGCATGATGATGCAGTTGTAAAGCTAAGCTGTCCGTTGGAACTCATCCAGTTTCA
AAGGTGATACAAACGTTGAGGGAGAATGAAGTTATGCCTCATGATTCCAACGTGGCCATC
ACAGAGGAGGGTGTGGTTCACACATTCACCTCTCCGGCCTCAGGGTGGCTGCACCGCTGAG
CAGTTGAAGGACAAGCTTCTGCCTCTCTATCACAGTAACATCACAGCAGTAACGTGCTA
TGTAATAAGTGTAACCGTGTGGAGGTTGTATCAATGTACTATTGCAAGCCAACCAAAAA
AACTCCAGCTTAGTAGGATCGTGAATTTTCTTATATGTAATGTTGAGATTTGTCTTT
TACATATAAAGATTTGA

>G588 Amino Acid Sequence (domain in AA coordinates: 309-376)
MGQKFENQEDRAMVESTIGSEACDFFISTASANTALSCLVSPSPSDSNLQQLRHVVEG
SDWDYALFWLASNVNSDGCVLWGDGHCVRVKKGASGEDYSQQDEIKRRVLRKLHLSFVG
SDEHRLVKGALTDLDMFYLALYFSFRCDTNKYGPAGTYVSGKPLWAADLPSCLSYR
VRSFLARSAGFQTVLSVNVSGVVELGSLRHIPEDKSVIEMVKS VFGGSDFVQAKEAPKI
FGRQLSLGGAKPRSMSINFPKTEDDTGFSLESYEVQAIGGSNQVYGYEQGKDETLYLTD
EQKPRKRGRKPANGREELNHVEAERQRREKLNQRFYALRAVVPNISKMDKASLLADAIT
YITDMQKKIRVYETEKQIMKRRESNQITPAEVDYQQRHDDAVVRLSCPLETHPVSKVIQT
LRENEVMPHDSNVAITEEGVVHTFTLRPQGGCTAEQLKDKLLASLSQ*

>G1758 (69..677)
GTCCCTCCTCTTAGCTTCAACCGCCGAAAAACTAAACAACCTTCTTGAAAAAAGAGA
AACTAAAAATGAACCTCTTCAACCCCTAACCTAGCTCCACAGATTTCACTGAATTTT
TCAAGTTTCGATGATTTTTCAGGATACTTTTGAGAAGATCATGGAAGAAATCGGCCGTGAGG
ACCACTCGTCGTACCGACTTTGAGTTGGAGTTCATCGGAAAAGTTAGTGGCTGCAGAAA
TCACAAGCCCCGCTTCAACAAGCCTAGCTACCTCACCCTATGAGCTTTGAAATAGGTGACA
AAGATGAAATCAAAAAGAGGAAGAGACACAAAGAAGATCCGATTATTCACGCTTCAAAA
CGAAATCATCAATTGATGAAAAGGTTGCTTTAGATGATGGGTATAAATGGAGGAAATACG
GAAAGAAGCCGATAACGGGTAGTCCATTTCCAAGGCATTATCACAAGTGTTTCGAGCCAG
ATTGCAACGTGAAGAAGAAGATCGAAAGAGATACGAACAATCCAGATTACATATTGACAA
CATACGAAGGTAGACATAACCAACCAAGCCCTTCTGTAGTTTATTGTGATTCAGACGACT
TTGATCTTAACCTCTCTCAACAATTGGTCTTTTCAGACGGCAAATACGTATAGTTTCTCTC
ATTCTGCTCCATATTGATCGATCGTAGTTACAAGTTTGTGTATATAGATGTATATATATA
TATACCAATTACCATCGTAATCACGTCTCACATGTAACGTACGTACATATATCTTGTTT
GGGGTTCGTTTGTAAATGTAATTGGTGGAGGTAGAATGGAAGTCATCTTGATAGT
TGTACTTGATGTAAGGTTTGATAGTCATTTTTTATAAAGTAACATAATTGTACAA

>G1758 Amino Acid Sequence (domain in AA coordinates: TBD)
MNYPSNPNSSTDFTEFFKFDDFDDTFEKIMEEIGREDHSSSPTLSWSSEKLVAAEITS
PLQTSLATSPMSFEIGDKDEIKKRKRHKEDPIIHVFKTKSSIDEKVALDDGYKWRKYGKK
PITGSPFPRHYHKCSSPCDNVKKIERDTNNPDYILTTYEGRHNHPSVSVYCDSDDFDL
NSLNNWSFQTANTYSFHSAPY*

>G2148 (66..737)
GTCTCTAATATAAGCTTGAACGTTGCTATATATAAATGTAAAGGCGAACGCATAAGAAAA
GAAAAATGGAGAATGAAGCTTTTGTAGATGGTGAATTGGAGTCTCTTTGGGGATGTTCA
ACTTTGATCAATGTTTCATCAACGAATCGAGCTTTTGCAATGCTCCAAATGAGACTGATG
TTTTCTCTCTGATGATTTCTTCCCATTTGGTACAATTCTGCAAAGTAACATGCGGCCG
TTCTTGATGGTTCCAACCAACCAAGCAACCGAAATGTCGACTCAAGACAAGATCTGTTGA
AACCAGGAAGAAGCAAAAGTTAAGCTCGGAAAGCAATTTGGTTACCGAGCCTAAGACTG
CTTGAGAGATGGTCAAAGCCTAAGCAGTTATAATAGTTTCAGATGATGAAAAGGCTTTAG
GTTTAGTGTCTAATACATCAAAAAGCCTAAAACGCAAAGCGAAAGCCAACAGAGGGATAG
CTTCCGATCCTCAGAGCCTATACGCTAGGAAACGAAGAGAAAGGATAAACGATAGGCTAA
AGACATTGCAGAGCCTAGTTCTTAATGGGACAAAGGTCGATATAAGCACAAATGCTGGAAG
ATGCTGTCCATTACGTGAAGTTCTGTCAGCTTCAAATCAAGCTCTTGAGTTCAGAAGATC

TATGGATGTATGCACCTCTTGCTCACAATGGTCTGAATATGGGACTACATCACAATCTTT
TGTCTCGGCTTATTTAAGACAAAATCATTGGAATAACATAACTTACAGTACTTGT'TTTT
TTCTCGTCTATATTCATGATTATGGTTATTTTTTTGTTTGAGTTGTTCAATTTTCTGTC
TATTGCGTCTATGAAC'TTGACACTCTTTTGTAAATTATTATATGCTAAAGACAATTGG
ACTAACAGCATT'TTAATAAAAAAAAAAAAA

>G2148 Amino Acid Sequence (conserved domain in AA coordinates:130-268)

MENEFVDGELESLLGMPNFDQCSSNESSFCNAPNETDVFSSDDFFPFGTILQSNYAAVL
DGSNHQTNRNVDNRQDLLKPRKKQKLSSSNLVTPEKTAWRDGGQSLSSYNSSDDEKALGL
VSNTSKSLKRKAKANRGIASDPQSLYARKRRERINDRLKTLQSLVPNGTKVDISTMLEDA
VHYVKFLQLQIKLLSSEDLWMYAPLAHNGLMGLHNNLSRLI*

>G2379 (52..798)

CGCCGCTCACTCTCCTCCCGGTGCCGCACATTAGCAACACTACTCCCGACGAATGGAGACG
ACGACGCCGCGAGTCAAAATCAAGTGTGTCCACCGACCGCCGTTGGGAAGAGAAGACTGG
TGGAGTGAGGAAGCGACGCGACGCTGGTAGAAGCCTGGGGCAATCGTTACGTCAAGCTG
AACCACGGAATCTCCGGCAGAATGACTGGAAAGACGTCGCCGACGCCGTTAACTCTAGA
CACGGTGATAACAGCCGTAAGAAGACCGACTTACAGTGTAAGAACCGGGTCGATACTTTG
AAGAAGAAGTACAAAACAGAGAAAGCTAAACTCTCGCCGTCGACTTGGCGTTTCTATAAC
CGCCTCGATGTTCTAATCGGTCCCCTGTGAAGAAATCGGCTGGCGGAGTTGTCAAATCA
GCGCCTTTTAAGAATCATCTGAATCCAACTGGATCGAACTCTACTGGAAGCTCTCTTGAA
GATGATGATGAGGATGATGATGAGGTTGGTGATTGGGAATTCGTTGCTAGGAAGCATCCT
CGTGTGGAAGAGGTAGATCTGAGTGAAGGATCAACGTGTAGGGAAGTACTACGGCGATT
CTCAAGTTTGGAGAAGTTTACGAGAGAATTGAAGGAAGAAGCAACAGATGATGATTGAG
TTGGAGAAGCAGAGAATGGAAGTGACAAAGGAGGTAGAGTTAAACGAATGAACATGTTG
ATGGAGATGCAGTTAGAGATTGAGAAATCAAAGCACCGGAAACCGCAAGTGCTTCAGGT
AAGAAGAACTCACATTAGG

>G2379 Amino Acid Sequence (domain in AA coordinates:19-110, 173-232)

METTTTQSKSSVSHRPPLGREDDWSEEATATLVEAWGNRYVKLNHGNLRQNDWKDVADAV
NSRHGDNRSRKKTDLQCKNRVDTLKKKYKTEKAKLSPSTWRFYNRLDVLIGPVVKKSAAGV
VKSAPFKNHLNPTGSNSTGSSLEDDDEDDDEVGDWEFVARKHPRVEEVDLSEGSTCRELA
TAILKFGEVYERIEGKKQMMIELEKQRMVETKEVELKRMNMLMEMQLEIEKSKHRKRAS
ASGKKNH*

>G1462 (63..1031)

CGTCGACCATCTTGCGGATTGATCTTCTCTAGATAATTTTTTTGATCGATTTAGTTTCA
TTATGGAGGACGACGACGCGAGCTTATGATCTAATCAAACACGAACTGTTATACTCAGAAG
ACGAAGTAATAATCTCACGTTATCTGAAGGGTATGGTCGTTAACGGAGATTCTTGGCCAG
ATCACTTCAATCGAAGACGCAACGTTGTTACCAAGAATCCAGATAAGGTGTTCAATTCTG
AGAGACCTAGATTCTGTGATCGTTAAACCACGAACAGAGGCTTGTGGTAAACCGATGGAT
GTGATTCCGGTTGCTGGAGGATCATTGGTCGTGATAAACTGATAAAGTCGGAGGAGACTG
GGAAGATTCTAGGTTCAAGAAGATACTCAAGTTTTCCTAAAGAGGAAACCTATAGACT
ACAAGAGAAGTTGGGTAATGGAAGAGTATAGGCTTACCAATAACTTGAAGTGAAGCAAG
ATCATGTGATTTGCAAAATTCGGTTTATGTTTGAAGCTGAAATTAGTTTCTTGCTAAGCA
AGCATTTCTACACTACATCAGAAATCGGTTCTTGAAAATGAGCTGTTGCCATCTTATGGAT
ATTATTTATCCAATACACAAGAGGAGGATGAATTTTATCTGGACGCGATAATGACTTCGG
AAGGAAACGAGTGCCCTAGCTACGTTACCAACAACGTGTACTGTCTGCATCCATTGGAGC
TTGTGGATCTTCAAGATCGGATGTTAATGATTACGGAACCTGCATCTTCGCTAACAAGA
CTTGTTGGTGAAACTGATAAATGCGATGGTGGTTACTGGAAGATCCTGCACGGTGATAAGC
TGATCAAGTCAAATTTGCGAAAGGTCATTGGTTTCAAGAAGGATTTTGAGTTCTATGAAA
CGGTGAGACAAATATATCTTTGTGATGGAGAAGAAGTGACCGTAACCTGGACTATACAAG
AGTATAGGCTTAGCAAAAACGTGAAGCAGAATAAAGTGTGTGCGTTATCAAGTTGACTT
ATGATAGATAGGATCTTTACTTTGGTTTTTGTGATCATCTTAGTATCTTACGAATATTC
TAGATACACACATCTATAGGCGACCGCTCTAGACAGGCTCGTACCG

>G1462 Amino Acid Sequence (domain in AA coordinates: TBD)

MEDDDAAYDLIKHELLYSEDEVIISRYLKGMMVNGDSWPDHFIEDANVFTKNPDKVFNSE
RPRFVIVKPRTEACGKTGDCDSGCWRIIGRDKLIKSEETGKILGPKKILKFLKRPIDY
KRSWMMEYRLTNLNLWKQDHVICKIRFMFEAEISFLLSKHFYTTSESVLNELLPSYGY
YLSNTQEEDEFYLDAMTSEGNEWPSYVTNNVYCLHPLELVDLQDRMFNDYGTICIFANKT
CGETDKCDGGYWKILHGDKLIKSNFQKVIQKVFYETVRQIYLCDGEEVTVTWTIQE

>G1211 (44..1120)

TGAAACCTAGATTCTGCAACTGAATTCCTAATTCGAAAAAGAAATGGAGGGTTCGTCGTC
GACGATAGCAAGGAAGACATGGGAAC TAGAGAACAGCATTTCTAACAGTAGACTCACCTGA
TTCAACCTCCGACAACATCTTCTACTACGACGATACTTCACAGATCTAGGTTCCAGCAAGA
GAAACCGTGGGAGAAATGATCTCTACTACTTTAAACAGTTCAGAGATCTCAGCGCTCGCTCT
TCTTAAGATGGTGGTTACGCTCTCGCTCTGGTGGTACAATTGAAATAATGGGTCTTATGCA
AGGTAAGACCGATGGTGATACTATCATTGTTATGGATGCTTTTGC'TTTACCAGTGGAAGG
TACTGAGACAAGGGTTAATGCTCAGGATGATGCTTATGAGTACATGGTTGAGTATTCACA
GACCAACAAGCTCGCGGGGCGGCTGGAGAAATGTTGTTGGATGGTATCACTCTCACCCTGG
ATATGGATGCTGGCTCTCCGGTATTGATGTTTCTACGCAGACGCTTAAACCAACAGCATCA
GGAGCCATTTTACGTCTGTTGTTATTGATCCCAAGGACTGTTTTCAGTCTGGTAAGGTTGA
GTTGGTGCTTTTCAAGAACATACTCTAAAGAGATATAAGCCTCCAGATGAACCTGTTTCTGA
GTATCAAACATATCTCTTTAAATAAGATTGAGGACTTTGGTGTTCAGTGCAAACAGTACTA
TTCATTAGATGTCACTTATTTCAAGTCATCTCTTGATTCTCACCTTCTGGATCTACTATG
GAACAAGTACGGGTGAACACTCTTTCTCTCTCCACTGCTGGGTAAATGGAGACTATGT
TGCTGCGACAAATATCAGACTCTAGCTGAGAAGCTTGAGCAAGCCGAGAGTCACTGTTTCA
GTCTCGCTTTGGAGGAGTTGTGCCATCACTCCCTTCATAAGAAAAAAGATGAGTCTCA
ACTTAACTAAGATAAATCGGGATACGCGAAAGATAAATCTGGAAACAGGTCATGGAC'TAAT
GTCGCAGGTCATAAAAGATGAATTATTCAACTCAATGCGTCAGTCCACAACAAATCTCC
CACTGACTCGTCGGATCCAGACCTATGATTACATATTGAAGTTGCTCTCTCTTTTGGTTT
CTANTTTTGGATTGACCCATCATTTGTTGTCTCTTCATTATTTTCTGTTGTGTAAAGAA
TTATAATGNCNGCNGCGAATTCGCGGCGCTAAAAAANACAGGAATTTGAAANAATTCN
NCCATTGCCAATCTTTTAAATTATATTATCTCTCNAATTATATAATTCAAACATCCCT
ANTANCTTCATTGACCGTCCCCCTCCCTCCCGTGTGTCNTGGTGCTGGCCCC

>G1211 Amino Acid Sequence (domain in AA coordinates: 123-179)

MEGSSSTIARKTWELENSILTVDSPDSTSDNIFYYDDTSQTRFQOEKPWENDPHYFKRVK
ISALALLKMVVHARSGETIEMGLMQGKTGDGTIIVMDAFALPVEGTETRVNAQDDAYEY
MVEYSQTNKLAGRLENVVGWYHSHPGYGCWLSGIDVSTQTLNQHQHEPFLAVVIDPTRTV
SAGKVEIGAFRTYSKYGVPPDEPVSEYQTIPLNKIESTDFGVHCKQYSSLDYTFYKSSLSHS
LLDLLWNKYWNTLSSPLLGNVDYVAGQISDLAEKLEQAESHLVQSRFGGVVPSLSLHKK
KEDESQLTKITRDSAKITVEQVHGLMSQVIKDELFSNMRQSNKSPSTDSSDDPMITY*

>G1048 (5..892)

GACCATGGCGGAGGAATTTGGAAGCATAGATTCTACTCGGAGATGAAGATTTCTTCTTCGA
TTTCGATCCTTCAATCGTAATTGATTCTCTTCCGGCGGAGGATTTCTTCAGTCTTCACC
GGATTCATGGATCGGAGAAATCGAGAATCAATTGATGAACGATGAGAATCATCAAGAGGA
GAGTTTGTGGAATTGGATCAGCAATCGGTTTCAGATTTTCATAGCGGATCTACTCGTTGA
TTATCCAATGACGATTCTGGCTCCGTTGATTGGCGGCTGATAAAGTCTAACCGTCGA
TTCTCCCGCCGCCGTGATGATTCCGGGAAGGAGAATTCGGAATTTGGTTGTTGAGAGA
GTCTAATGATTCTGGTAGCGAGATTTCATGATGATGACGAAGAAGGAGACGATGATGC
TGTGGCTAAAAACGAAGAAGGAGAGTAAGAAATAGAGATGCGGCGGTTAGATCGAGAGA
GAGGAAGAAGGAATATGTACAAGATTTAGAGAAGAAGAGTAAGTATCTCGAAAGAGAATG
CTTGAGACTAGGACGTATGCTTGTAGTGCTTCGTTGCTGAAAACCAAGTCTCTACGTTACTG
TTTGCAAAGGGTATGCGTAATGACTACCATGATGTCGAAGCAGGAGTCTGCTGTGCT
CTTGTTGGAATCCCTGCTGTTGGGTTCCCTGCTTTGCTCTGGGAGTAAACTTCATTTG
CCTATTCCCTTATATGTCCACACAAAGTGTTGGCTCCTACGTCCAGAACCAGAAAAGCT
GGTCTAAACGGGCTCGGGAGTAGTAGCAACCGTCTTATACCGGCGTTAGTCGGAGATG
TAAGGGTTCGAGGCCTAGGATGAAATACCAAATCTTAACCCCTTGGCGGTGACAACGCCT
TTTTTAACGTCTCTTTTTCGCGCATTTTGGATTGTAGATGAGTGCTTTTAGTTTTCTCTC
TCTTGTTTGTATTTGCTGTTGAAAGTTTCTGTCTAATATCGATAAGTTAACAGTGAA
AAAAAAAAAAAAAAAA

>G1048 Amino Acid Sequence (domain in AA coordinates 138-190)

MAEEFGSIDLLGDEDFFFDFDPSIVIDSLPAEDFLQSSPDSWGIEIENQLMNDENHQEES
FVELDQQSVSDFIADLLVDYPTSDSGSVDLAADKVLTVDSPAAADD SGKENS DLVVEKKS
NDSGSEIHDDDEEGDDDAVAKKRRRRVNRMDAAVRSRERKKEVYVDGLEKKSKYLNERECL
LGRMLCEFVAENQSLDRYLQKGNNGNTTMRSDQESAVRLLESLLGLSLLLWLVNFICL
FPYMSHTKCCLLRPEPEKVLVNLGLSSSKPSYTGVSRRCKGSPRMKYOILTAA*

>G986 (31..846)

CATTAAATTGGCTCCTGTGAACCTAAATTTATGGACTATGATCCCAACACCAATCCGTTTC
GACCTTCATTTCTCCGGTAAACTTCCGAAAAGAGAAGTCTCGGCTTCAGCTTCTAAAGTT
GTAGAGAAGAAATGGTTAGTGAAAGATGAGAAGAGAAATATGCTACAAGATGAAATAAAC
CGGGTTAATTCGGGAGAACAGAAGCTAACCGAAATGTTAGCAAGAGTCTGTGAGAAGTAC
TATGCTCTTAATAATCTTATGGAGGAGTTGCAGAGTCGAAAGAGTCCTGAAAGTGTTAAC
TTTCAGAACAAACAGCTAACGGGGAAACGAAAACAAGAACCTTGATGAGTTTGTAGCTCC
CCAATTGGACTCAGTCTCGGACCAATCGAGAACATCACCAACGATAAAGCGACGGTTTCA
ACCGCTTACTTTGCTGCTGAGAAGTCTGACACAAGCTTGACTGTGAAAGATGGATATCAA
TGGAGGAAATACGGGCAAAGATTACGAGAGATAATCCATCTCCTAGAGCTTACTTCAGA
TGCTCGTTTTTCACCGTCTTGCTTAGTCAAGAAGAAGGTGCAACGAAGTGCAGAAGATCCA
TCTTTCTTGGTAGCCACTTACGAAGGGACACATAACACACCCGACCACATGCAAGTGTG
TCCAGGACAGTGAAACTTGATCTAGTTCAAGGTGGGCTTGAACCAAGTTGAGGAAAAGAAA
GAGAGAGGGACGATTCAAGAGGTTTGGTGCAACAAATGGCTTCTTCGTTGACCAAAGAT
CCTAAGTTCACTCAGCTCTTCGCACTGCTATTTCCGGGAGATTGATAGAGCATTCAGA
ACATGAAAGTTCTCTAGAACATGTATATTTCTGTTTTGTTCTATTTTGTGCTCATTCCCT
AGTAAAAAGGTAAAGATTTGTTTGATCTTGATTAGGAGGCATAGATGTCAATTTTAATGT
GTGTGTATATAATTACATCAAACTAAGTATCCAAAAGGGTCACCCCCATTTTATCTTA
TG

>G986 Amino Acid Sequence {domain in AA coordinates: 146-203}
MDYDPNTNPFDLHPSGKLPKREVSASASKVVEKKWLVKDEKRNMLQDEINRVNSENKKLT
EMLARVCEKYYALNMLMEELQSRKSPESVNFQNKQLTGKRKQELDEFVSSPIGLSLGPIE
NITNDKATVSTAYFAAEKSDTSLTVKDG YQWRKYGQKITRDNPSPRAYFRCSFSPSCLVK
KKVQRSAEDPSFLVATYEGTHNHTGPHASVSRTVKLDLVQGGLEPVEKKERGTIQEVLV
QQMASSLT KDPKFTAALATAISGR LIEHSRT*

>G789 (259..1593)

GGCAAGAAGAACCTTAGCCTCTCTTTCTTCTTCTCTCTCTCTCTCTGTGGTACTGTT
CTGTTTCAACTTTACTCCCTCAGTTTTCAGAACAAATCCCTATCTAGAAGAGAGATAAAAC
CGAGAAGGTTTTGGAGATAGAAATCTTTTGTCTTCTTTTGTCCCTCCTTGCTCGATTTTT
GTTACGTGTGAAGCAATAAAAAAACTGATATAGCTAAATCTTCCATCCATTCAGAGGC
TTCTAAATCTGATCTGACATGGAACAAGTGTGCTGATTGGAATTTTGAAGATAATTTT
CACATGTCCACTAATAAAAGATCAATCAGACCAGAAGATGAATTAGTGGAGCTATTGTGG
AGAGATGGTCAAGTGGTTTTACAAAGCCAAAGCTCGTAGAGAACCGTCAGTCCAAGTCCAA
ACCCACAAACAAGAAACCTTAAGAAAACCAACAATATTTTCTTGACAACCAAGAAACA
GTACAAAAGCCTTACTACGCTGCTCTAGATGATCAAGAAACCGTCTCCTGGATACAATAC
CCTCCGGATGACGTCATCGACCCCTTTCGAATCCGAGTTCTCCTCTCATTCTTCTCTTCTG
ATCGATCACCTCGGAGGTCCTGAGAAGCCACGAACGATCGAAGAGACAGTTAAGCATGAG
GCTCAAGCCATGGCTCCTCCTAAGTTTAGATCCTCGGTTATAACAGTCCGACCGAGTCAT
TGCGGCAGCAACCAAGTCAACAAATATTTCATCAGGCCACTACACTTCCGGTTTCTATGAGT
GATAGAAGCAAGAAGCTCGAAGAAAGACTTGACACTTCGTCAGGTGGCTCCTCCGGTTGC
AGCTATGGAAGGAACAACAAGAAACCGTTAGTGGAAACAGTGTAACCAATTGACCGTAA
AGAAAACATGTTATGGATGCTGATCAAGAATCTGTGTCTCAATCAGATATAGGTTTGACC
TCAACCGATGATCAAAACCATGGGTAACAAATCGAGCCAACGGTCAGGATCTACTCGAAGA
AGCCGTGCAGCTGAAGTTCATAATCTCTCAGAAAGGAGGAGAGATCGGATCAATGAA
AGAATGAAAGCTCTTCAAGAACTCATACCTCACTGCAGCAGAACAGATAAAGCTTCGATA
TTGGATGAAGCAATTGATTACTTAAAATCACTTCAAATGCAACTCCAAGTGATGTGGATG
GGAAGTGGAAATGGCGGCGGCGGAGCAGCAGCAGCAAGTCCGATGATGTTTCCCGGGGTA
CAATCATCTCCATACATTAATCAGATGGCTATGCAAAGTCAGATGCAATTGTCTCAATTC
CCGGTTATGAACCGGTCCGCTCCGAGAACCATCCCGGTTTAGTATGTCAAACCCGGTA
CAGTTGCAGCTCCAAGCACAGAACAAATCTTATCGGAGCAGCTCGCTAGGTACATGGGC
GGGATTCCCCAGATGCCGCCGGCGGGAATCAGATGCAGACCGTGCAACAACAACAGCG
GACATGTTGGGATTTGGATCTCCGGCGGGACCGCAAAGTCAACTGTCCGCCACCGGCGACC
ACCGACAGTCTTCATATGGGTAAAATAGGCTGACTTGGCATATAGTTTTCTCCGAAATT
ATTCTTCTTACAGTTGGTGATGTTATTTATTTTGGTGCCTAAGCAAGCATAAAAGCT
AAGTCAAATGTATTATAGAGATCTAATAAGTTAGTCTCATACTTATAACTTATTTTTAA
CAGTTGAATTATAGTATCAATCAAGTGTGGGAACCTAAAGATCATACATGTGTCAATAC
TTTTATATTTGTTCTCAAGTTTCATCAGAAAAACAAAATAAAAAGGATAGACTAGGCCTG

CATTGACATTATCATGGGCTTTTTTGGGTCTATGAATATGAACATTAACCCC *

>G789 Amino Acid Sequence (domain in AA coordinates: 253-313)
MEQVFADWNFEDNFHMSTNKRSIRPEDELVELLWRDQVVLQSQARREPSVQVQTHKQET
LRKPNNIFLDNQETVQKPNYAALDDQETVSWIQYPPDDVIDPFSEFSSSHFFSSIDHLGG
PEKPRTIEETVKHEAQAMAPPKFRSSVITVGP SHCGSNQSTNIHQATTLPVMSDRSKNV
EERLDTSSGGSSGCSYGRNNKETVSGTSVTIDRKRKHVMDADQESVSQSDIGLTSTDDQT
MGNKSSQSRSGSTRRSRAAEVHNLSERRRRDRINERMKALQELIPHCSRTDKASILDEAID
YLKSLQMQLQVMWMSGMAAAAAAASPMFPGVQSSPYINQMAMQSQMQLSQFPVMNRS
APQNHPLVLCQNPVQLQLQAQNLSEQLARYMGGIPQMPPAGNQMQTVQQQPADMLGFG
SPAGPQSQLSAPATTTDSLHMGKIG*

>G2085 (1..930)
ATGTTTGGTCGCCATTTCGATTATCCCAAATAACCAGATTGGTACCGCCTCTGCTTCCGCT
GGTGAAGACCATGTCTCTGCTCCGCTACGCTCTGGTCACATTCTTACGACGATATGGAA
GAAATCCCTCATCTGACTCTATCTATGGTGCTGCCTCCGATTTGATTCCCGATGGCTCT
CAATTGGTTGCTCACCGATCCGATGGCTCTGAATTACTTGTCTCTCGGCCACCGGAAGGG
GCGAATCAGCTTACGATCTCGTTCCGTGGACAAGTTTACGTTTTGATGCCGTTGGTGCT
GACAAGGTGGATGCTGTGTGTGTCGCTGTTGGGTGGTTCTACTGAGCTTGCTCCTGGTCCG
CAGGTGATGGAACTAGCTCAACAGCAGAATCATATGCCTGTTGTAGAATATCAGAGCCGC
TGTAGCCTTCCGCAACGGGGCACAATCCTTGGATAGGTTTCGGAAGAAGAGGAATGCTAGA
TGTTTTCGAGAAGAAAGTAAGATACGGGTGTTCCGCCAAGAAGTTGCCTTAAGAATGGCACGT
AATAAAGGTCAATTCACCTCTTCAAAGATGACAGATGGGGCTTATAACTCTGGCACAGAT
CAAGATTCTGCCCCAAGATGATGCCCATCCAGAAATATCGTGTA CTCTGCGGCATTAGT
TCCAAATGTACACCAATGATGCGACGTGGCCCTTCCGGCCCCAGGACTCTCTGCAATGCC
TGTGGACTTTTTTGGGCTAACAGGGGTACATTGAGGGATCTCTCAAAGAAAACAGAAGAG
AATCAGTTGGCTTTAATGAAACCGGATGATGGTGGGAGTGTGCTGATGCTGCTAACAAC
TTAAACACTGAAGCTGCAAGTGTGAAGAACAACACTTCCATGGTTTTCTCTTGCCAATGGG
GATAATTCTAATCTGTTAGGTGATCACTAA

>G2085 Amino Acid Sequence (domain in AA coordinates: TBD)
MFRHSIIPNNQIGTASASAGEDHVSASATSGHIPYDDMEIIPHPDSIYGAASDLIPDGS
QLVAHRSDGSELLVSRPPEGANQLTISFRGQVYVFDVAGADKVDVAVLSLLGGSTELAPGP
QVMELAQQQNHMPVVEYQSRCSLPQRAQSLDRFRKKRNARCFEKKVRYGVRQEVRLMAR
NKGQFTSSKMTDGA YNSGTDQDSAQDDAHPEISCTHCGISSKCTPMRRRGPSGPRTL CNA
CGLFWANRGTLRDLRSKKTEENQLALMKPDDGGSVADAANNLNTAASVEEHTSMVSLANG
DNSNLLGDH*

>G1783 (1..603)
ATGGCCCGCTTTCGCGAGTGGACAAGGGTCGATGACAAACGTTTGTAGTTAGCTCTGCTT
CAAATCCCGGAGGGTTCGCCGAATTTTATAGAGAATATCGCCTATTATCTCCAGAAACCG
GTGAAGGAGGTGGAGTACTACTGCGCGTTGGTCCATGATATTGAGCGGATCGAATCG
GGTAAGTATGTTTGGCCAAATACCCGGAAGACGATTACGTGAACTGACGGAAGCAGGT
GAGTCTAAGGGCAATGGGAAAAAGACGGGAATTCCTTGGTCAGAAGAGGAACAGAGGTTG
TTTCTGGAAGGACTAAATAAGTTTGGGAAAGGAGACTGGAAGAACATATCGAGGTATTGT
GTGAAGTCAAGGACCTCGACGCAAGTGGCAAGCCATGCTCAGAAGTATTTTGAAGGCAA
AAGCAGGAGAGTACGAATACTAAACGCCCGAGTATTCATGACATGACTCTGGGAGTTGCG
GTCAATGTCCCTGGATCCAACCTTGAGTCTACTGGCCAGCAACCACATTTTGGTGATCAA
ATTCCTTCGAATCAATATTATCCCTCCCGAGGAAAACCTTTCGGGGTTTTGATCAGCGATGG
TGA

>G1783 Amino Acid Sequence (domain in AA coordinates: 81..129)
MAAFPQWTRVDDKRFELALLQIPEGSPNFIENIAYYLQKPVKEVEYYCALVHDIERIES
GKYVLPKYPEDDYVKLTEAGESKGNKKTGIPWSEEEQRLFLEGLNKFQKGDWKNISRYC
VKSRTSTQVASHAQKYFARQKQESTNTRKPSIHDMTLGAVVNVPGSNLESTGQQPHFGDQ
IPSNQYYPSQENFRGFDQRW*

>G2072 (155..793)
TCGACCCACGCGTCCGCCACGCGTCCGATCTTTTCACAGAAGACCAACCAGCTTGGCT
CGATGAGCTCCTAAGTGAGCCAGCATCACCTAAGATTAACAAAGGTCATAGACGTT CAGC
TAGTGACACAGCTGCTTACTTGAACCTAGCTTTAATGCCTTCGAAGGAAAATCATGTTGC
TGGTTCTGCTTGGCAGTTCAGAACTATGATTTGTGGCAGTCCAACCTCTTATGAACAACA
CAATAAATTAGGATGGGATTTCTCTACAGCAAATGGAACATAATATCCAAAGAAATATGTC

ATGCCGAGCTTTAAATATGTCGTCGAAACCCATTGAGAAACATGTAAGCAAAATGAAAGA
AGGAACTTCTACAAAACCAGATGGTCCTAGATCAAAGACTGACTCAAAACGTATCAAACA
TCAAAATGCTCATCGAGCGCGTTTGAGAAGGCTTGAGTACATATCAGACCTTGAAAGGAC
CATCCAAGTGCTACAAGTTGAAGGATGTGAAATGTCATCTGCCATTCACTACTTGGATCA
GCAGTTACTCATGCTTAGCATGGAAAATAGAGCTTTAAACAACGTATGGATAGTTTAGC
AGAAATCCAAAAGCTTAAACATGTGGAGCAGCAATTGCTTGAGAGAGAGATAGGAAACCT
ACAGTTTTCGACGACCAACAACAACACAGCAAAACCAAAACAAGTCCAAGCAATACA
AAATCGATACACCAATATCAACCACCTGTTACACAAGAACCCGATGCCAATTTGCAGC
CTTGGCAATATGATTAGGAAATATGGATACATTGTTTCAGATTAAGCTGAGCTCCTCTTG
CTCTACCTTAATGTCCATACAACATAGGTGAACCTGATGTTTGTAGCCTTGAATGAAAAC
CTAAAAAAGCATCGTTATGTAAATCAAAATGTGGTTGCCCATATCCTCCTCTATTGCATT
TCTCTCTATTTATGGCATGGTAGAGAATCTCTTGTCAGAAACCTCATGTTATGTAATAA
CTGTTAATCCTCTTATTTTCATCTATTATATATATGAATAAGTAATTTTTTTGCCAAAAA
AAAAAAAAAAAAAAAAAAAA

>G2072 Amino Acid Sequence (conserved domain in AA coordinates:90-149)

MPSKENHVAGSSWQFQNYDLWQNSYBQHNKLWDFSTANGTNIQRNMSCGALNMSSKPI
EKHVS KMKEGTSTKPDGPRSKTDSKRIKHQNAHRRLRRLEYISDLERTIQVLQVEGCEM
SSAIHYLDQQLMLSMENRALKQRMDSLAEIQKLKHVEQQLLREIGNLQFRRHQQPQO
NQKQVQAIQNRKYQPPVTQEPDAQFAALAI*

>G931 (85..1071)

GGAGGTTCTTTGACAGACACATGTATCATCAATCTTCTCTGTTGAAGCAGAGAGAGAG
AGCTAATTGTTGCCCTCGAGTCACATGGATAAGAAAGTTTCATTACTAGCTCTGTGGCA
CATTCAACTCCACCATACTTAGTACTTCCATCTCATGGGACTTCCAACCAATCCAAT
GGTGTGACTGAATCACTGAGTTTGAAGGTGGTAGATGCAAGACCAGAACGCTTTATAAAC
ACAAAGAAATATCAGTTTCCAGGACCAGGATTCATCTTCAACTCTGTCTCTGCTCAATCT
TCTAACGATGTTACAAGTAGTGGAGATGATAACCCCTCAAGACAAATCTCATTTTATGCA
CATTGAGATGTTTGAAGGATTGAAGAACTCAAAGGAAGCGATTGCAATTAAATCA
GGCTCCTCCACGGCAGGAATCGCTGATATTCACTCTTCTCCTTCCAAGGCTAAGTTCTCA
TTTCACTATGCCGATCCACATTTTGGTGGTTAATGCCTGCGGCTTACCTACCACAGGCA
ACAATATGGAATCCCCAAATGACTCGAGTTCGCTACCATTGATCTCATAGAGAATGAG
CCTGTCTTTGTCAATGCAAAGCAATTCCATGCAATTATGAGGAGGAGGCAACAGCGTGCT
AAGCTAGAGGCGCAAAACAACTAATCAAAGCCCGTAAGCCGTATCTTCATGAATCTCGA
CATGTTACGCTCTTAAACGACCTAGAGGATCTGGTGAAGATTCTTAAACACCAAAAAG
CTTCAAGAATCTACAGATCCAAAACAAGACATGCCAATCCAACAGCAACACGCAACGGGA
AACATGTCAAGATTGTGCTTTTATCAGTTGCGAAGACAGCAATGACTGTGATTGTTCAACC
ACTTCTCGCTCTGACATCACATCTGCTTCTGACAGCGTTAATCTCTTTGGACACTCTGAA
TTTCTGATATCAGATTGCCCATCTCAGACAAACCAACAATGTATGTTTCATGGTCAATCA
AATGACATGCGATGGAGGTAGGAACACACACCATTTCTGTTCATATCTGAGCCGTTGGA
ATCTGGTAAATGTGTACGTTTCTACAAAAAAGGGGAAGTCATCCTTGGCTGCTACTTCGCT
TATTAGCTAGTTCTTATTTACACGCTTTGTCCAGATATC

>G931 Amino Acid Sequence (domain in AA coordinates: TBD)

MDKKVSFTSSVAHSTPPYLSTSIWGLPTKSNVTELSLKVVDARPERLINTKNISFQD
QDSSSTLSSAQSSNDVTSSGDDNPSRQISFLAHSVDVCKGFEETQRKRFAIKSGSSTAGIA
DIHSSPSKANFSFHYADPHFGGLMPAAYLPQATIWNPMQTRVPLPFDLIENEPVFNKQ
FHAIMRRRQQRKLEAQNKLIKARKPYLHESRHVHALKRPGRSGGRFLNTKKLQESTDPK
QDMPIQQQHATGNMSRFVLYQLQNSNDCDCSTTSRSDITSASDSVNLFGHSEFLISDCPS
QTNPTMYVHGQSNDMHGGRNTHHFSVHI*

>G278 (93..1874)

TCGATCTTTAACCAAAATCCAGTTGATAAGGTCTCTTCGTTGATTAGCAGAGATCTCTTTA
ATTTGTGAATTTCAATTCATCGGAACCTGTTGATGGACACCACCATGATGGATTCCGCCG
ATTCTTATGAAATCAGCAGCACTAGTTTCGTGCTACCGATAACACCGACTCCTCTATTG
TTTATCTGGCCGCCGAACAAGTACTACCGGACCTGATGTATCTGCTCTGCAATTGCTCT
CCAACAGCTTCGAATCCGTCTTTGACTCGCCGGATGATTTCTACAGCGACGCTAAGCTTG
TTCTCTCCGACGCGCCGGAAGTTTCTTCCACCGGTGCGTTTTGTGTCAGCGAGAAGCTCTT
TCTTCAAGAGCGCTTTAGCCGCCGCTAAGAAGGAGAAAGACTCCAACAACACCGCCGCCG
TGAAGCTCGAGCTTAAAGGAGATTGCCAAGGATTACGAAGTCGGTTTCGATTTCGGTTGTGA
CTGTTTTGGCTTATGTTTACAGCAGCAGAGTGAGACCGCCGCCCTAAAGGAGTTTCTGAAT

CGCGAGACGAGAATTGCTGCCACGTGGCTTGCCGGCCGGCGGTGGATTTCATGTTGGAGG
TTCTCTATTTGGCTTTTCATCTTCAAGATCCCTGAATTAATTACTCTCTATCAGAGGCACT
TATTGGACGTTGTAGACAAAGTTGTTATAGAGGACACATTGGTTATACTCAAGCTTGCTA
ATATATGTGGTAAAGCTTGTATGAAGCTATTGGATAGATGTAAAGAGATTATTGTCAAGT
CTAATGTAGATATGGTTAGTCTTGAAAAGTCATTGCCGGAAGAGCTTGTTAAAGAGATAA
TTGATAGACGTAAAGAGCTTGGTTTGGAGGTACCTAAAGTAAAGAAACATGTCCTCGAATG
TACATAAGGCACTTGACTCGGATGATATTGAGTTAGTCAAGTTGCTTTTGAAGAGGATC
ACACCAATCTAGATGATGCGTGTGCTCTTCATTTGCTGTGTCATATTGCAATGTGAAGA
CCGCAACAGATCTTTTAAACCTTGATCTTGCCGATGTCAACCATAGGAATCCGAGGGGAT
ATACGGTGCTTTCATGTTGCTGCGGATGCGGAAGGAGCCACAATTGATACTATCTCTATTGG
AAAAAGGTGCAAGTGCATCAGAAGCAACTTTGGAAGGTAGAACCGCACTCATGATCGCAA
AACAGCCACTATGGCGGTTGAATGTAATAATATCCCGGAGCAATGCAAGCATTCTCTCA
AAGGCCGACTATGTGTAGAAATAC TAGAGCAAGAGACAAACGAGAACAAATTCCTAGAG
ATGTTCCCTCCCTCTTTTGCAGTGGCGGCGGATGAATTGAAGATGACGCTGCTCGATCTTG
AAAATAGAGTTGCACCTTGCTCAACGCTCTTTTCCAAACGGAAGCACAAAGCTGCAATGGAGA
TCGCCGAAATGAAGGGAACATGTGAGTTTCATAGTACTAGCCTCGAGCCTGACCGTCTCA
CTGGTACGAAGAGAACATCACCGGTTGTAAAGATAGCACCTTTCAGAATCCTAGAAGAGC
ATCAAAGTAGACTAAAGCGCTTTCTAAAACCGTGGAACCTCGGGAACGATTCTTCCCGC
GCTGTTCCGCGAGTGTCTGACCAAGATTATGAAGTGTGAGGACTTGACTCAACTGGCTTGCG
GAGAAGACGACACTGCTGAGAAACGACTACAAAAGAAGCAAAGGTACATGGAAATACAAG
AGACACTAAAGAAGGCCCTTTAGTGAAGACAATTTGGAATTAGGAAATTCGTCCTGACAG
ATTCGACTTCTTCCACATCGAAATCAACCGTGGAAAGAGGTCTAACCGTAAACTCTCTC
ATCGTCGTCGCTGAGACTCTTGCTCTTAGTGTAATTTTGTGTACCATATAATTCTGT
TTTCATGATGACTGTAACTGTTTATGTCTATCGTTGGCGTCATATAGTTTCGCTCTTCGT
TTTGTCATCCTGTGTATTATTGCTGCAGGTGTGCTTCAAACAAATGTTGTAACAATTTGAA
CCAATGGTATACAGATTGTAAATATATATTTATGTACATCAACAATAAAAAAAAAAAAAA
AAAA

>G278 Amino Acid Sequence (domain in AA coordinates: 2-593)
MDTTIDGFADSYEISSTSFVATDNTDSSIVYLAEEQVLTGPDVLSALQLLSNSFESVFDSP
DDFYSDAKLVLSDGREVSFHRCLVSARSSFFKSALAAKKEKDSNNNTAAVKLELKEIAKD
YEVGFDSVVTVLAYVYSSRVPPPKGVSECADENCHVACRPVDFMLEVLYLAFIPKIP
ELITLYQRHLDDVVDKVVIEDTLVILKLANICGKACMKLLDRCKEIIVKSNDVMVSLEKS
LPEELVKEIIDRRKELGLEVPKVKHVS NVHKALDSDDI ELVKLLKEDHTNLD DACALH
FAVAYCNVKTATD LLLKLDLADVNHRNPRGYTVLHVAAMRKEPQLILS LLEKGASASEATL
EGRTALMIAKQATMAVECNNIPEQCKHSLKGRLCVEILEQEDKREQIPRDVPPSFAVAAD
ELKMTLLDLENRVALAQRLFPTQAAMEIAEMKGTCEFI VTSLEPDRLTGKTRTSPGVK
IAPFRILEEHQSRKALSKTVELGKRFFPRCSAVLDQIMNCEDLTQLACGEDDTAEKRLQ
KKQRYMBIQETLKKAFSEDNLELGNSSLTDSSTSSKSTGGKRSNRKLSHRRR*

>G2421 (1..630)
ATGGAGGGTTCGTCCAAAGGGTTGAGGAAAGGTGCATGGACTGCTGAAGAAGATAGTCTC
TTGAGGCAGTGTATTGGTAAGTATGGAGAAGGCAAATGGCATCAAGTTCCTTTAAGAGCT
GGGCTAAATCGGTGCAGGAAAAGTTGTAGACTAAGATGGTTAAACTATTTGAAGCCAAGT
ATCAAGAGAGGAAAATTTAGTTCTGTATGAAGTTGATCTTCTTCTTCGTCTTCATAAGCTT
CTAGGAAATAGGTGGTCTCTGATTGCTGGTCGATTACCTGGTCGGACCGCTAATGATGTC
AAGAACTACTGGAACACCCATCTGAGTAAGAAGCATGAACCGTGTGTAAACTAAGATA
AAAAGGATAAATATTATAACCCCTCCTAATACACCGGCCCAAAAAGTTTGTGAAAATAGT
ATCACATGTAACAAAGATGATGAGAAAGATGATTTTGTGGATAATTTTATGGTTGGAGAT
AATATATGGTTGGAGCGTTTGTCTAGACGAGGGCCAAGAGGTAGATGTGCTGGTTACAGAA
GCGGCGGCAACAGAAAAGGAGGGGCACTTTGGCGTTTGACGTTGAGCAACTTTGGAATTTG
TTCGATGGAGAGACTGTGATCTTTGATTAGTGTATTATAAACGTTTGTGTTCTCTGTTTG
TGAGGTTTCTCTATTTAATTTAGTATCTATTTTCTAAATTAACATAATATCTTATAGTATT
TTAGGCAAAACCTTATGTTTCCGTTTCTGTCGCGCCGCTCTAG

>G2421 Amino Acid Sequence (domain in AA coordinates: 9-110)
MEGSSKGLRKGAWEEDSLLRQCIGKYGEGKWHQVPLRAGLNRCKRKSRLRWLNLYLKPS
IKRGKFSSEVDLLRLHLKLLGNRWSLIAGRLPGRANDVKNYWNTHLSKKHEPCCKTKI
KRINIITPNTPAQKVCENSITCNKDEKDDFVDNFMVGDNIWLERLLDEGQEVLDVLVTE
AAATEKEGTLAFDVEQLWNLFDGETVIFD*

>G2032 (53..1789)

TCCTCCACAGTAGTAAGAACTTCCATACTTTGCTCTAGATTTCCTTGAGAAAAGATGCAGCC
GATCTTCCATGCGATCCTTAAAAATGACCTTCCAGCTTTTTTAGAGTTGGTAGAAGATAG
TGAATCGTCTCTGGAGGAGAGAAACGAGGAAGAACTTGAACAAACACGGTTTTGCACAT
GGCTGCAAAGTTTGGTCACCGAGAACTCGTCTCCAAGATTATTAGACTCCGACCTTCCCT
CGTGTCTTCCCGCAACGCATACAGAAACACCTTTGCATCTTGCTGTATCTCTGGAGA
TGTAACCATAGTTATGTCAGATGTTAGAGACTGGATTGGAAGTGTGTTCTGCGACGCAATAT
CAACAACCATACACCACTTCCACTTGGCTTGCCGTAGCAATTCCATAGAGGCTGCCAGACT
CATCGCGGAAAAGACACAATCAATTGGCCTCGGTGAACTCATTTCTGCCATATCAAGTGG
ATCCACTAGTATCGTAGGGACTATACTGGAGAGATTCCAGACCTAGCTAGGGAAGAAGC
TTGGGTGTTTGAAGACGGCTCACAACTAACGCTACTGCATCATCGTGTGATAAGGGGAGA
CTTTGAACGTGACAACATATATTGTAGGGCTCGATCAAGGATTAGAAGAAGCACTTAACCC
CAATGGTTTATCACCCTTGATCTCTGGCTCCTCAGAGCTCGGTGTGATCTCTGGAGGA
GTTCTTGGACAAGGTTCCATTGTCTTTTACGCTCAATTACGCCGTCGAAAGAGACAGTCTT
TCATCTCGCTGCTCGAAACAAAAATATGGATGCCTTTGTTTTTATGGCAGAGAGTTTGGG
AATTAACAGCCAAATTCTTCTACAGCAAACCGATGAAAGTGGCAACACTGTCTTACATAT
TGCTGCATCCGTCTCTTTTGATGCTCCTCTTATACGTTACATTGTTGGTAAGAATATAGT
AGATATCACGTCGCAAGAACAAAGATGGGTTTTGAAGCTTTTCAACTTCTCCCTCGAGAAGC
CAAGACTTTGAGTTGTTTATCAAGTGGCTGAGATTGGTACCGACCTTCCAGAGAGCT
GGAATCTGAGAACAAATGTAGAAACAACACGAAGGCTCTCAAGAGGTGAGGTAATACGGTT
GCTAAGGATTATAGGAATAAACACATCAGAGATAGCAGAGAGAAAGAGAAGCAAGGAACA
GGAAGTGGAAAGAGGTCTGCAGAACTTGGAAATATCAGATGCATATAGAAGCATTACAGAA
TGCAAGAAATACGATTGCTATAGTGGCAGTCTTGATTGCTTCAGTTGCTTATGCCGGTGG
GATAAACCCCTCCGGGGGGCGTCTACCAAGACGGCCATGGAGAGGGAAATCCTTAGTGGG
GAAAACAACGGCGTTTTAAGTCTTTTGGCATATGCAACAACATCGCATGTTCCAGTCCTT
GGGCATCGTTATTCTTTTGTGTAGCATATACCTTACAAGAGGAAACCCCTTAAGAGGATT
ATTGGTGGCCACGCATAGGATGATGTGGGTTTCTGTAGGTTTCATGGCGACGGCTTATAT
AGCGGCGCTTTGGGTGACCATACCGCATTATCATGGAACACAATGGTTATTTCCAGCAAT
TGTAAGCCGTTGCTGGTGGAGCGTTGACCGTACTCTTTTCTATCTCGGAGTTGAGACCAT
CGGTTCATTGTTTAAAGAGATGAATCGTGTAGGGGATAATATACCTTCCCTTTGCAAGAAC
CAGTTTCAGATTTAGCCGCTTCCGGAATAACAGGCTATTTACCTATTAAGAAAACTGGT
TTTTCTAATTTCCCTGTAACCTGTGTAAATTGTGTATGTG

```
>G2032 Amino Acid Sequence (domain in AA coordinates: entire protein)
```

MQPIFHAILKNDLPAPFLELVEDSESSLEERNEEEHLNNTVLHMAAKFGHRELVSKIIELR
PSLVSSRNAYRNTPLHLAAILGDVNIVMQMLETGLEVCSARNINNHTPLHLACRSNSIEA
ARLIAEKTQSIGLGELILAISSGSTSIVGTILERFPDLAREEAWVVEDGSQSTLLHHACD
KGFELTTILLGLDQGLEEALNPNGLSPLHLAVLRGSVVILEEFLDKVPLSFSSITPSKE
TVFHLAARNKNMDAFVFMAESLGINSQILLQQTDESGNTVLHIAASVSFDAPLIRYIVGK
NIVDITSKNKMGFEAFQLLPREAQDFELLSRWLRFGTETSQELDSENNVEQHEGSQEV
IRLLRIIGINTSEIAERKRSKEQEVERGRONLEYQMHI EALQNARNTIAIVAVLIASVAY
AGGINPPGGVYQDGPWRGKATSLVGKTTAFKVFAICNNIALFTSLGIVILLVSIIPYKRKPL
KRLLVATHRRMMVQSVGFMAITYAASWVTI PHYHGQTQWLFPAIVAVAGGALTVLFFYLVG
ETIGHWFKKNRVGDNIPISFARTSSDVLAVSGKSGYFTY*

>G1396 (83..313)

TCGACCTCGTTTCCTTTCCCTCCTCTCTTCTACCATTAGTACGTTACTGGAGCTGATCTC
ACGTATATTTTGGATCGTAATCATGGACGGCGAAGATTTTGCCGGAAAGGCGGCTGCTGA
AGCCAAGGGATTGAACCCGGGATTAATCGTGCTGCTTGTGTGGAGGTCCGCTTCTTGT
GTTCTTAATCGCCAAC TACGTGCTTTACGTTTATGCTCAGAAGAACCTACCTCCAAGGAA
GAAGAAGCCCGTTTCCAAAAAGAAGCTCAAGCGGGAGAAGCTAAAGCAAGGAGTCCCTGT
CCCTGGAGAATAAAAGCCAGCTTAAGCTTCCCTTCACTTGTGCCTCCTTCAAAGCGGTTTT
TGTCGGTTACCAAATTCACCCCTTGCGGGTTTTTTCTTCTCTTACTTCTGT CATGAGG
ATTATCTTTTGAGGCCTT

>G1396 Amino Acid Sequence (domain in AA coordinates: TBD)

MDGEDFAGKAAAEAKGLNPGLIVLLVVGPLLVLFLIANYVLYVYAQKNLPPRKKKPVSKK
KLKREKLKQGVPPVGE*

>G619 (382..2748)

ATTTTTTCCAATCTGCAAATTTAGTCTATGTCTGTTCCCTTGCTCCCTCTTCTCAGT

ACCTGCAAAATGGAGGAAGAAGAAATCCTTCTCTGAAACCCCTTGTTCTCATTGATTCCTCC
TTCCTCTCTCTTCTCTCTCTCTCTGATTCTGTTATTCCACACTTATGACTCATCTT
TCCCGTCAATAGCTAAGTTTGCCTCTTCTTTGTGAAATTTAGCTGAAAAAGGAGAGGAAT
TCCGAATTCTGTCACTTCAAAGCTCGAATTTTGCAAACTTTCTTTGATGGGTTTTACTT
GTTTTGTGTGAATCTGATTAAAAATAGAACTTTTGTCTTCTTGTCTCCTTTTGCT
CTTAAAGAGAAGCTTTTCAATGGAATTTGACTTGAATACTGAGATTGCGGAGGTGGAA
GAGGAGGAGAATGATGATGTAGGAGTAGGAGTAGGAGGAGGAACAAGAATTGACAAGGGT
AGGCTTGGAAATTTACCATCTTCTTCTTCTCATGCTCTTCCGGATCATCATCGTCATCA
TCTTCTACAGGCTCTGCATCTTCCATTTACTCTGAGCTTTGGCATGCTTGTGCTGGTCCT
CTCATTGTCTTCCCAAGAAAGGCAATGTAGTTGTCTATTTCCTCAAGGTCATTTGGAG
CAAGATGCTATGGTTTCATATTCTCTCTCTTGAATCCCCAAATTTGACCTTAATCCC
CAAATCGTCTGCAGGGTGGTTAATGTCCAGTTGCTTGCTAATAAGGACACCGATGAGGTC
TACACTCAAGTCACTCTGCTTCCACTTCAAGAGTTTTCGATGCTAAATGGGGAGGGGAAA
GAGGTCAAGGAGTTAGGAGGGGAGGAAGAGAGGAACGGAAGCTCATCCGTCAAGCGGACA
CCTCATATGTTCTGTAAAACCTTAACAGCGCTGTACACAAGCACACATGGAGGCTTCTCT
GTACCTAGAAGAGCCGCTGAAGATTGTTTTGCTCCTCTTGACTACAAACAACAGAGGCCA
TCTCAAGAGCTCATTTGCAAGGACCTCCATGGAGTAGAGTGGAAGTTTCGCCATATCTAT
AGAGGTCAACCAAGGAGCATCTACTCACCACCTGGTTGGAGTATCTTTGTCACTCAAAAG
AATCTCGTCTCTGGTGATGCGGTTCTTCTTGAGAGACGAAGGAGGAGAGCTGAGATTA
GGAATCAGAAGAGCAGCACGGCCAAAGAAATGGACTTCTGACTCAATCATTGAGAAGAAT
TCATGTTCAAACATTTCTGTCTCTTGTGGCTAATGCTGTATCTACAAAAGCATGTTTCAT
GTGTTCTACAGTCCACGAGCGCAGCATGCAGAGTTTGTGATTCTTATGAGAAGTATATC
ACAAGCATCAGGAGTCTGTTTGCATAGGCACAAGATTAGAATGCGATTGAAATGGAC
GATTCTCTGAGAGAAGATGCGCTGGTGTAGTGAAGTCTGTGACTTGGACCCGTAT
AGGTGGCCAAACTCTAAATGGAGGTGCTTGTGGTGCGATGGGATGAGTCTTTTGTGAGT
GATCATCAAGAAAGAGTTTACCTTGGGAGATTGATCCCTCGGTTTCTCTCCACACTTG
AGCATTCAGTCATCTCCAAGGCTTAAAGGCCATGGGCAGGTTTACTGGATACTACCCCA
CCCGGAAACCCCATAAACAAAAGGGGTGGTTTTTTTGGACTTTGAGGAGTCGGTTAGACCC
TCTAAGGCTTGTCAAGGTCAAGAAAATATAGGTTCTGCATCACCCCTCACAGGGGTTTGAT
GTTATGAACCGCCGATACTGGATTGTCGATGCAGTCTCATGCAATCCAGTCTTGTG
TCGAGTAGAGTCAAGGATCGATTGTTGGTGAGTTTGTAGATGCTACTGGCGTGAACCCAGCT
TGTTTCAGGTGTTATGGACCTGGATAGGTTTCCAAGGGTCTTGCAAGGTCAAGAAAATTTGC
TCGCTTAAATCATTCCCGCAATTTGCTGGTTTTCAGTCCAGCTGCTGCTCCTAATCCCTTT
GCTTACCAAGCCAACAAGTCAAGTTACTATCCGCTAGCTTTGCATGGGATTAGGAGCACT
CATGTTCCGTATCAGAATCCATACAAATGCGGGAAACCAATCCTCGGGTCCCCCTTCACGT
GCAATAAACTTTGGTGAAGAGACTAGAAAAGTTTGATGCACAAAATGAAGGTGGCCTACCA
AATAATGTTACAGCTGATTGTCATTTCAAGATTGATATGATGGGAAAACAGAAAGGCAGT
GAGTTGAATATGAATGCTTCAATCAGGATGTAACTTTTCGGATTCTCCTTACCAGTGGAG
ACACCTGCATCTAAGCCGCAAGCTCGAGCAAAAGAATCTGTACAAAGGTTTACAAGCAA
GGAAGCCAAGTGGGGAGAGCTATTGATTGTGCGGACTTAACGGGTATGATGATCTCCTT
ATGGAGCTTGAACGGCTGTTCAACATGGAAGGGCTTCTCAGGGATCTGAAAAAGGATGG
AGGATCTTATATACTGATAGTGAACGATATGATGGTCTGGCGATGATCCATGGCAT
GATTTCTGCAATGTGGTGTGGAAGATACACTTATACACGAAAGAGGAAGTGGAGAATGCG
AATGACGATAACAAGAGTTGTTTAGAGCAAGCTGCTCTCATGATGGAAGCATCAAAGTCA
TCTTCTGTGAGCCAGCCTGATTCTTCTCTTACAATCACTAGGGTTTGATACCCATAAAGA
AGCTTATTTCTATGTTTTAAAGTGTGTTTGTCTCACAAAAGAACTTCAACTTTATCTTT
GTCTTTGAATCCATTTATGTGTTTGTGTTTCTTCTGGTCTCCATGGATGTCTCATG
TGTACCGTTTTACTGAGAGATATGTGAGTTTATGGGATGTGTAAAGCATGCCATTGGAT
TTTAAGGTTTTCAAATTACAATATATATATTAGTTTTGAAGTTAAAAA

A
>G619 Amino Acid Sequence (domain in AA coordinates: 64-406)
MEFDLNTIEIAVEVEEENDVDVGVGGGTRIDKRLGISPSSSSSSSSSSSSSSSTGSAS
SIYSELWHACAGPLTCLPKGNVVVYFPQGHLEQDAMVSYSSPLEIPKFDLNPQIVCRVV
NVQLLANKDTDEVYTVQVLLPLQEFMSLNGEGKEVKELGGEERNGSSSVKRTPHMFCKT
LTASDTSHTGGFSVPRRAEDCFAPLDYKQQRPSQELIAKDLHGVEWKFRHIYRGQPRRH
LLTTGWSIFVSQKNLVSGDAVLFLRDEGGELRLGIRRAARPRNGLPDSII EKNSCSNLS
LVANAVSTKSMFHVFSYPRATHAEFVIPYKYITSIRSPVICGTRFRMRPFEMDDSPERRC

AGVVTGVCIDLDPYRWPNSKWRCLLVWRWDESFVSDHQERVSPWEIDPSVSLPHLSIQSSPR
PKRPWAGLLDTPPGNPITKRGGFLDFEESVRPSKVLQGOENIGSASPSQGFDMNRRIL
DFAMQSHANPVLVSSRVKDRFGEFVDATGVNPACSGVMDLDRFPRVLQGOEICSLKSFQ
FAGFSPAAPNPFAYQANKSSYYPLALHGIRSTHVYPQNPYNAGNQSSGPPSRAINFGEE
TRKFDAQNEGGLPNNVTADLPFKIDMMGKQKGSELNMNASSGCKLFGFSLPVETPASKPQ
SSSKRICTKVHKQGSQVGRAIDLRLNGYDDLMELERLFNMEGLLRDPEKGWRILYTDS
ENDMMVVGDDPWHDFCNVVKIHLTYKEEVENANDDNKSCLEQAALMMEASKSSSVSQPD
SSPTITRV*

>G2295 (33..917)

GTAATATAACAATAACTCAGGTTACAAAGGATGGTTCGGAAAGTGGTCGACCTACAAA
GGATAGCGAACGATAAGACAAGGATAACAACCTTACAAGAAGAGGAAAGCTAGTCTTTTACA
AGAAGGCACAAGAGTTCTCAACTCTCTGCGGCGTCGAGACATGTCTCATCGTCTACGGTC
CCACGAAGGCTACCGATGTGGTGATTTCCGAGCCAGAGATATGGCCGAAGGACGAGACCA
AAGTCAGGGCCATCATACGCAAGTACAAAGACACAGTGTGACCCAGCTGCAGGAAAGAAA
CCAACGTGGAGACTTTCGTCAACGATGTAGGGAAGGAAACGAGGTGGTGACTAAAAAGA
GAGTGAAGCGTGAGAATAAGTATTCTAGTTGGGAGGAGAAGCTAGACAAGTGTTCACGAG
AGCAACTACATGGGATTTTCTGTGCGGTGGATAGCAAGTTAAATGAAGCTGTAACGAGAC
AGGAGCGTAGTATGTTTAGGGTTAATCATCAAGCCATGGACACACCATTCCCGCAGAATT
TAATGGACCAACAATTTCATGCCACAGTATTTTCATGAGCAGCCACAGTTTCAAGGCTTCC
CTAATAATTTCAATAATATGGGTTTCTCGTTGATTTACCTCATGATGGTCAGATTCAAA
TGGACCCCAATCTCATGGGAGAGTGGACCGACTTGGCTTTGACTCAAAGCTTGATGATGT
CAAAGGGAACGATGGTACTCAATTTCATGCAGAGGCAAGAACAACCATATAATCGTG
AACAGGTTGTATCGAGGTCTGCAGGTTTCAATGTAAACCGTTTATGGGATATCAAGTCC
CGTTTAATATCTCTAATTGGAGATTATCGGGAATCAAGTTGAAAATTGGGAGCTTTCAG
GGAAGAAAACGATATGATTTGAATTACGGAGCTTTATTAGTTTTTAGGGTTTTATAGTTT
TG

>G2295 Amino Acid Sequence (domain in AA coordinates: TBD)

MVPKVVDLQRIANDKTRITTYKKRKASLYKKAQEFSTLCGVETCLIVYGPTKATDVVISE
PEIWPKDETKVRAIRKYKDTVSTSCRKETNVETFVNDVGKNEVVTKKRVKRENYSSW
EEKLDKCSREQLHGIFCAVDSKLNBAVTRQERSMFRVNHQAMDTPFPQNLMDQQFMPQYF
HEQPQFQGFNNFNNMGFSLISPHDGGIOMDPNLMKWTDLALTQSLMMSKNGDGTQFMQ
RQEOPYYNREQVVSRSAGFNVNPFMGYQVFNIPNWRLSGNQVENWELSGKTTI*

>G312 (1..1755)

ATGGCTTACATGTGCACTGATAGTGGCAATCTAATGGCTATTGCTCAACAAGTCATCAAA
CAGAAGCAGCAACAAGAACAACAACAGCAGCAACATCATCAAGACCATCAGATTTTTGGT
ATTAATCCTTTGTCTCTTAACCCATGGCCCAATACTTCCCTCGGGTTTGGGCTTTCAGGT
TCGGCTTTTCCCGACCCGTTTCAAGTTACCGGCGGCGGAGATTCCAACGATCCTGGCTTT
CCTTTTCTTAACCTTAGACCACCAACACGCAACAACCCGCGGTGGGTTCAAGTTATCT
GATTTCCGGCGGTGGAACCGCGCGGCGGAGTTTGAGTCCGACGAGTGGATGGAGACTCTT
ATCAGCGGTGGAGACTCCGTTGCAGACGGTCCTGATTGTGACACCTGGCATGATAATCCC
GATTACGTAATCTACGGTCTCTGATCCATTTCGATACTTACCGAGTCGACTCAGTGTCCAA
CCGTGAGATCTAAACCGAGTCATTGACACGTCGAGTCCGCTTCCCTCCGCGACCTTGTGG
CCTCCTTCTTCGCCATTATCGATTCTCCGCTTACTCATGAGTCACCAACCAAAGAAGAT
CCAGAGACTAACGACTCCGAAGACGATGACTTCGACCTAGAACCACTCTCCTCAAAGCT
ATATACGACTGTGCACGGATCTCAGACTCTGACCCTAACGAAGCTTCCAAGACGCTTCTT
CAGATCCGAGAATCTGTATCGGAGCTAGGTGATCCGACGAGCGAGTTGCATTTTACTTC
ACGGAAGCTCTCTCCAACAGACTGTCTCCTAATTCCGCCGCGACGTCGTCCTTCTTCTCA
TCTACGGAGGATTTAATCTTATCTTATAAAACCTAAACGACGCTTGTCTTACTCCAAA
TTGCGACATTTGACGGCGAATCAAGCGATTCTAGAAGCGACGAGAGAAGTCGAACAAGATT
CACATCGTCGATTTTGAATCGTTCAAGGTATACAATGGCCTGCTCTTCTTCAAGCTCTA
GCTACTCGTACTTCTGGTAAACCCACTCAAATCCGGGTCTCGGGTATACCCGCTCCATCT
CTCGGTGAATCTCCGGAACCGTCGTTAATCGCCACCGGAAACCGCTCCGTGATTTCCGCC
AAGGTTCTGGATCTGAATTTGATTTTCATCCCAATTCTCACTCCCATACATTTACTTAAC
GGGTCAAGTTTCCGGGTGACCCCGGATGAAGTACTGGCCGTGAATTTTCATGCTCCAGCTC
TACAATACTCGACGACGCGGAGATAGTTGACACCGCACTACGGCTCGCCAAATCG
TTGAACCCGAGGGTCGTCACCTCTCGGAGAATACGAAGTGAGCTTAAACCGGGTCCGGTTTC
GCTAACCGGGTAAAGAACGCGCTTCAATTCTATTCCGCGGTTTTCGAATCCCTTGAACCG

AACTTGGGGCGTGATTCCGGAGGAGAGAGTGAGAGTTGAGCGAGAGTTGTTCCGCCGAGAGA
ATCTCGGGTTTGATTGGACCGGAGAAAACCGGAATTCATAGAGAAAGAATGGAAGAGAAA
GAGCAATGGCGGGTATTAATGGAGAATGCCGGTTTGAATCGGTTAAGCTGAGTAATTAC
GCAGTGAGCCAAGCGAAGATTCTATTGTGGAATTACAATTACAGCAATTTGTATCAATT
GTTGAATCTAAGCCTGGCTTCATCTCTTTGGCCTGGAACGATTACCTCTCCTCACTCTT
TCTTCTCGCGATAA

>G312 Amino Acid Sequence (domain in AA coordinates: 320-336)
MAYMCTDSGNLMAIAQQVIKQKQQEQQQQHHQDHIQIFGINPLSLNPWPNTSLGFGLSG
SAFPDPFQVTGGGDSNDPGFPFNNLDHHHATTTGGGFRLSDFGGGTGGGEFESDEWMETL
ISGGDSVADGPDCTDWHNDPDIYIGPDPFDTPSRLSVQPSDLNRVIDTSSPLPPPTLW
PPSSPLSIPPLTHESPTKEDPETNDSDDDFDLEPPLLKAIYDCARISDSDPNEASKTLL
QIRESVSELGDPTERVAFYFTEALSNRLSPNSPATSSSSSTEDLILSYKTLNDACPYSK
FAHLTANQAILEATEKSNKIHI VDFGIVQGIQWPALLQALATRTSGKPTQIRVSGIPAPS
LGESPEPSLIATGNRLRDFAKVLDLNFDFIPILTPIHLLNGSSFRVDPDEVLA VNFMLQL
YKLLDETPTIVDTALRLAKSLNPRVVTLGEYEVSLNRVGFANRVKNALQFYSAVFESLEP
NLGRDSEERVVRVERELFGRRIISGLIGPEKGTIHRERMEEKEQWRVLMENAGFESVKLSNY
AVSQAKILLWNYNYSNLYSIVESKPGFISLAWNDLPLLTSSWR*

>G1444 (192..1001)
AATCCCCATATCCTTCGCAAGACCCCTTCTCTATATAAGGAAGTTCATTTTATTTT
GACACGCTGACAAGCTGACTCTAGCATATCTGGCACCGGCGACCACTCTTTTGGTGC
AAAGATCCCAAAAATCAAAATCGAAAGAGAGAATAAATCAAAAGGAAGAATCTTTATCT
GCTTTCTCTCGATGAGGATCCGGAAACGACAAGTGCCTCTTCTTTATCGTCTCTATTAC
CAGTTCCTCTATCAGATCTCTACTTTAACCCTCACCGACGCGCCACCGCGAGATACTTTC
CGGGTGGTTATAAAGACGGCGGTGATGATTTTGGTTCTCTTCAGCTTTTCGCTTCCGCCGC
CGTCGAGATTTCTGATCGGCTTATTCAAAGAGATTGATAAAGAAGAAGGAGGAGGTCA
AGGCTTTGGATGATGATAATGGTGATGTAGACGTCAAGAGTCGTACTGATGCATCGGGCA
GCAAGAATGTTAATCCCCGAGGAGAATCCGTCTCTTCAATACAAGTTGTCGAGAAGAATG
AAAAGGTTGTGCTTTGAGGAAGAGAAGAGGCTTTATCAACTTTGAGGATTACGAAGATG
AGGAAGATGAAGAAGCTAGTGGCGGTGGAGGCCGTATTAATAAGGGAAAAAGAAAGCGA
AAAAGAGCGGTGGTGGGTAGAGGAAGGATCACGGTGCAGCCGTGTTAACGGTAGAGGAT
GGAGATGTTGTGACAAACGCTTGTGGTTATTCTCTTTGTGAGCATCATCTCGGTAAAG
GAAGGGTAAGGAGCTGAACAAGAGTGGTGGTGGTGGTGGCGCGGAGAAAAAGGCGGTGG
TGGTGAAGTGAAGAAGAAGAGAGTAAAGCTTGGCATGGTAAAGGCACGTTCAATAAGTA
GTTTGCTTGGACAAACCAGCACTAGTGGTGGTACTAGTGGTGATGTTGATCAGGGTGAGA
TAAGTGCACCTGCTGATCAGTTCGCTGCATGTGATAAGTAGGTCTGTTGATCAGCATTTG
CATGTATATGGATATGTGTATGTTTATGTACATGATGATAATGGGCATAGCGCGGCCGCT
CTAGACAGGCCTGGAACCGGATCTCTAGCTAGAGCTTTCGTTAGTATCATCGGGTTTAG
ACAACGTT

>G1444 Amino Acid Sequence (domain in AA coordinates: 168-193)
MRIRKQVPLPLSSLLPVPLSDLYFNRSPTATARYFRGGYKGGDDFGSLQLSLPPPSQI
SDRLIQRDLIKKKEEVKALDDNDGVDVKSRTDASGSKNVNPRGESVSSIQVVEKNEKV
SLRKRRGFINFEDYEDEEASGGGGRINKGKKKAKKSGGGLGEGSRCSRNVNRRGWRC
QQTLVGYSLCEHHLGKGRVRS MNKSGGGRGGEKKAVVVEVKKRVKLGVMKARSISLLG
QTSTSGGTSGDVDQGEISAPADQFAACDK*

>G801 (27..746)
GATAGTGATAACGAAATCCTAATTCATGGCCGACAACGACGGAGCAGTGAGTAACGGCA
TCATAGTCGAGCAGACGTCAAAACAAAGGACCTCTTAACGCCGTTAAGAAACCAACGCTTA
AAGATCGACACAGCAAGTTGACGGAAGAGGAAGAAGGATTTCGTATGCCAATCATTTGCG
CAGCTCGAGTTTTTCAATTGACCAGAGAGTTAGGTCAAGTCCGATGGTCAAACCATAG
AGTGGCTTCTCCGTCAAGCTGAGCCTTCTATCATAGCCGCCACTGGAACCTGGCACTACTC
CGGCGAGTTTCTCCACTGCTTCTCTCTCCACTTCTTCTCCGTTTACTCTCGGGAAACGTG
TCGTGAGAGCGGAGGAAGGAGAATCCGGCGCGGAGGAGGAGGGTTAACAGTGGGAC
ACACAATGGGGACTTCGTTAATGGGTGGTGGTGGTCTGTTGGTGGTGGTGGTGGTGGTGGT
CGAGGCCGATTTTCGACAAAGTCTGGAGCTTTGCAACCGGAGCTCCACCGGAAATGGTTT
TTGCGCAGCAGCAGCAACAGCTACACTTCTCGTCCGCCACCAGCAGCAACAGCAAGCTT
CCGCCCGCGCAGCAGCTGAATGGGTGAGGCTTCAGCAGCTAGAGTTGGGAATTATCTTC
CGGGTCATCATCTCAATTTGCTTGCTTCTTGTCTGGTGAGCTAACGGGTCCGGTCCGGA

GGGAAGACGACCACGAACCACGTTGAGAAATGGTATTGTCTTTTGGTAATGTATAGAAA
AATTCCTATGTTTTATGTCATCGAAAGTGTTAGAAAGTACCTCTAATTTGCGGTTTCTT
TTGCTCCTTTTTTACTTAATTTAAGCTTATTGCTTGTGTTGATTAGGGTTTTAGGGTTTAA
GAATATTTGGTCTCGTTAATTTGTTTCGGAGAGTGATAGAAAGAGAGAGATTGATTGA
TTGTTGTACCTAAAACGCTATAAAAAGCTCTGTTTTTACTAGCGAAAAAA

>G801 Amino Acid Sequence (domain in AA coordinates: 32-93)
MADNDGAVSNGIIVEQTSNKGPLNAVKKPPSKDRHSKVDGRGRRIRMPIICAAARVFQLTR
ELGHKSDGQTIIEWLLRQAEPSIIAATGTGTTPASFSTASLSTSSPFTLGKRVVRAEEGES
GGGGGGGLTVGHTMGTSLSLGGGGSGGFVAVPARPDFGQVVSFATGAPPEMVFAQQQQPAT
LFVRHQQQQASAAAAAAMGEASAARVGNYPGHHLNLLASLSGGANGSGRREDDHEPR*

>G1950 (42..764)
CTGAATTCGAACTTTGGAGAAGAAGAAGCTTTGATCAATCATGGAAATTGCAACCGATA
CAGCAAAGCAGATGAGAGACGAAGAGTTGTTCAAAGCAGCGGAATGGGGAGATTTCATCGT
TGTTTCATGTCTATTATCTGAAGAACAGCTCTCTAAATCTCTCAATTTAGAAACGAAGATG
GTGCTCTCTCTCTCTCGATGCGTTCCTTCCTTCGCGCCATTCTCAAATAGTGAAGTTGTTAT
CAAGTTTCAGATGAAGCAAAGACTGTAATCAATAGCAAGGATGATGAAGGATGGGCTCCTT
TGCATTCCGCTGCTAGCATCGGTAATGCTGAGCTCGTTGAGGTGCTTTTGACCAGAGGTG
CTGATGTCAATGCCAAAAATAACGGTGGTTCGCTGCTCTTCACTATGCTGCTAGCAAAG
GCCGTTTGGAGATTGCTCAGCTTTTATTAAACACACGGTGCAAAGATTAAACATCACAGACA
AGGTTGGTTGCACTCCGCTTCACAGGGCAGCAAGCGTGGGAAAGTTAGAAGTTTGTGAAT
TTCTTATTGAAGAAGGAGCAGAGATCGATGCTACGGATAAAATGGGTCAAACCTGCACTCA
TGCAATTCAGTTATCTGCGATGACAAACAGGTTGCGTTCTCTGCTTATAAGACATGGTGCG
ATGTGGATGTAGAAGACAAGGAAGGCTACACTGTTCTAGGCCGAGCTACCAATGAATTCC
GACCTGCACTTATCGATGCTGCTAAGGCCATGCTTGAAGGATAAAATGACTCTGGATTAC
TTTAAACTTACTAATCTGAGAGTTGTTTAGTTACTTAAAGGATTTTTCTTTACTGTA
TCATGTTTGCAAATGTTTCTGCTTATCAATTCATGTTCTGT

>G1950 Amino Acid Sequence (domain in AA coordinates: 65-228)
MEIATDTAKQMRDEELFKAAEWGDSSLFMSLSEELSKSLNFRNEDGRSLHVAASFHGS
MEIVKLLSSSDEAKTVINSKDDGWAPLHSAASIGNAELVEVLLTRGADVNAKNNGGRTAL
HYAASKRLEIAQLLLTHGAKINITDKVGCTPLHRAASVGKLEVEFLIEEGAEIDATDK
MGQ TALMH SVICDDKQVAFLLIRHGADV DVEDKEGYTVLGRATNEFRPALIDAAKAMLEG
*

>G958 (55..1950)
CGTCGACATGTTTCATATTTGTTTCTAGCTAAGAAGTTTGTATAAGGCAGTGGACATGGCT
CCTGTTTCAATGCCCTCCAGGTTTCCGGTTTCATCCAACAGACGAAGAGCTTGTATATAC
TACCTCAAGCGAAAGATTAATGGTCGGACTATTGAGTTAGAGATAAATACCCGAGATTGAT
CTTTACAAATGCGAACCTTGGGATTACCTGGGAAGTCCTTGCTGCCAAGTAAAGACCTA
GAATGGTTCTTTTTTCACTCCTCGAGACCGGAAATATCCAAACGGATCAAGAACAAACCGG
GCGACCAAAGCAGGTTACTGGAAAGCCACCGGAAAGATCGTAAAGTGACTTCACATTCA
CGGATGGTTGGAACAAAGAAACATTAGTTTATTACCGAGGAAGAGCGCCTCATGGCTCT
CGTACCGATTGGGTCTATGCACGAGTACCGTCTTGAAGAACAAGATGTGACTCTAAATCC
GGTATACAGGATGCCATATGCACCTTGTTCGAGTATTTAAGAAGAGTGCTTTAGCCAACAAA
ATTGAAGAACCAACACCATGGTACGAAGAAGAACAAAGGAACGACTAATAGTGAACAATCT
ACTTCTAGTACTTGTTTGTATTCTGATGGAATGTATGAAAACCTCGAAAACCTCGGGGTAT
CCAGTCTCACCTGAGACAGGAGGCTTAACTCAACTCGGTAATAATTGTCGTCGGATATG
GAAACGATAGAGAATAAATGGAGTCAGTTTATGTCGCATGACACGTCCTTCAACTTCCCA
CCTCAGTCTCAATATGGAACAATCTCATATCCTCCTCGAAGGTTGATATAGCGTTAGAG
TGTGCAAGACTACAAATCGTATGTTGCCACCAGTACCACCACCTTACGTAGAAGGTCTC
ACACACAATGAATATTTTGGAAACAATGTAGCTAACGATACAGATGAAATGTTGAGCAAG
ATTATAGCATTTGGCTCAAGCCTCAGATGAGCCACGAACAGTCTAGACTCATGGGACGGT
GGTTCGCTTCCGGGAACCTTCATGGAGACTTTAACTATTCCGGAGAAAAAGTCTCATGC
CTAGAGGCGAACGTGGAGGCTGTAGATATGCAAGAACACCATGTGAATTTTAAAGGAAGAA
AGACTTGTGAAAACCTTGAGATGGGTAGGAGTATCAAGCAAGGAACCTGAAAAGAGCTTC
GTTGAAGAACAACCTCAACGGTAATTCCTATAGAAGATATTTGGAGATATCATAATGATAAT
CAAGAACAAGAACATCATGATCAAGATGGTATGGACGTTAACAACAACAATGGAGATGTG
GATGATGCTTTTCACTCGAGTTTTCGGAAAACGAACATAACGAGAATCTTTTGGACAAG
AACGATCATGAGACAACGAGTTCCTCATGTTTTGAGGTGGTAAAAAAGTTGAGGTTAGC

CATGGATTGTTTGT CACAAC TCGTCAGGTAACCAACACATTCTTCCAACAGATAGTACCA
TCGCAAACCGTTATAGTTTATATAAATCCGACGGATGGCAATGAGTGTTCATAGTATG
ACATCAAAAGAGGAGGTTTCATGTCCGTAAAAAGATAAATCCGCGAATCAACGGAGTAAGC
TCAACAGTTCTTGGACAATGGAGAAAATTCGCGCATGTTATTGGCTTCATTCTATGCTT
CTATTGATGCGTTGTGTTTCATCGAGGTAACCTAACAACAAACAGAGGCAGTGAAGGTTAC
TCGAGGCAGCCTACGAGAGGAGATTGTAACAATCGGGGAACAATACTCATGATGGAATAAT
GCTGTCTGTGAGAAGAAAATTTGGAAGAAGAAGAGAAAATATGGTTGACGAACAA
GGTTTTTCGTTTCAAGATAGTTTCGTATTGAAGAAGTTGGGGCTTTCTCTTGCTATCATC
TTAGCTGTTTCTACCATAAGTCTTATTGAATACTGAGGTTCAATATATCATATATGGCT
TTTCACTTTTCTATTGTACTCCCATTGCGCTAGGTCGTATGC

>G958 Amino Acid Sequence (conserved domain in AA coordinates: 7-156)

MAPVSMPPGFRFHPTDEELVIYYLKRKINGRTIELEIPEIDLYKCEPWLDPGKSLPSK
DLEWFFFSRDRKYPNGSRNRTKAGYWKATGKDRKVTSHSRMVGTKKTLVYYRGRAPH
GSRTDWMHEYRLEEQECDSKSGIQDAYALCRVFKKSALANKIEEQHHGTTKKNKGTNSE
QSTSSTCLYSDGMYENLENSGYPVSPETGGLTQLGNNSSSDMETIENKWSQFMSHDTSFN
FPPQSQYGTISYPPSKVDIALECARLQNRMLPPVPPLYVEGLTHNEYFGNNVANDTDEML
SKIIALAQASHEPRNSLSDWDGGSASGNFHDGDFNYSGEKVSLEANVEAVDMQEHVNFK
EERLVENLRWVGVSSEKELEKSFVEEHSTVPIPIEDIWRYHNDNQEQEHHDQDGMVDVNNNG
DVDDAFTLEFSENEHNENLLDKNDHETTSSSCFEVVKKEVSHGLFVTTROVTNTFFQOI
VPSQTVIVYINPTDGNECCHSMTSKEEVHVRKKINPRINGVSSSTVLGQWRKFAHVIGFIP
MLLLMRCVHRGNSNKNRSGEGYSRQPTRGDCNNRGITILMMENAVVRRKIWKKKKEKNMVD
EQGFRFQDSFVLKKLGLSLAILAVSTISLI*

>G1037 (1..1722)

ATGACTGTTGAACAAAATTTAGAAGCTTTGGATCAGTTTCCTGTAGGAATGAGAGTTCTT
GCTGTTGATGATGACCAAACCTTGTCTCAAAATCCTTGAATCTCTCCTTCGTCACTGCCAA
TACCATGTAACAACGACGACCAAGCACAAAAGGCTTTAGAGTTATTGAGAGAGAACAAG
AACAAGTTTGATCTGGTTATTAGTGATGTTGACATGCGCTGACATGGATGGTTTCAAATC
CTTGAGCTTGTGGTCTTGAAATGGACCTACCTGTCATAATGTTGTCTGCGCATAGTGAT
CCAAAGTATGTGATGAAGGGAGTTACTCATGGTGCTTGTGATTATCTACTGAAGCCGGTT
CGTATTGAGGAGTTGAAGAACATATGGCAACATGTCGTGAGAAGTAGATTTGATAAGAAC
CGTGGGAGTAATAATAATGGTGATAAGAGAGATGGATCAGGTAATGAAGGTGTTGGGAAT
TCTGATCCGAACAATGGGAAAGGTAATAGAAAACGTAAAGATCAGTATAATGAAGATGAG
GATGAGGATAGAGATGATAATGATGATTCTGTGCTCAAAAGAAGCAACGTGTTGTTTG
ACTGTTGAGCTGCATAAGAAATTTGTGAGCTGTTAACCAGTTGGGATATGAGAAGGCT
ATGCCATAAAAGATTTTGGATCTGATGAATGTTGAGAAGCTCACTAGAGAAAATGTGGCC
AGTCATCTTCAGAAAATTCGCGCTTACTTGAAGAGGATCAGTGGTGTTGGCTAATCAGCAA
GCTATTATGGCAAACCTGAGTTACATTTTATGCAAATGAATGGACTTGATGGTTTCCAT
CACCGCCCAATCCCTGTTGGATCTGGTCAGTACCATGGTGGGGCTCCTGCAATGAGATCT
TTCCCTTCAAACCGGATTTCTTGCGAGACTCAATAGCTCTTCGGGGATCGGTGTCCGCAGC
CTTTCTTCTCCTCCTGCAAGGAATGTTCTTGCAAAACAGACCGATATCGGAAAGTTTAC
CATGTCTCATCACTTCTCTTAACCACAGTGATGGAGGAAACATACTTCAAGGGTTGCCA
ATGCCTTTAGAGTTGACACAGCTTCAGACAAACAACAACAAAGTAGAAACATGAACAGT
AACAAGAGCATTGCTGGGACCTCCATGGCTTTCTTAGCTTCTCTACGCAACAAAACCTCG
CTCATCAGTGCTCCTAATAACAATGTCGTGGTTCTAGAAGGTCACCCACAAGCAACTCCT
CCAGGCTTCCCAGGACACAGATCAATAAACGTTTGGAGCATTGGTCAAATGCTGTATCC
TCTTCGACTCACCTCTCTCCCGGCACATAACAGTAATAGTATCAATCATCAGTTTCGAT
GTCTCTCCATTACCGCATTCTAGACCCGACCCCTTGGAATGGAACAATGTGTATCAAGC
TACTCTATAACATTCTGTGACTCTGCCAATACATTGAGTTCTCCAGCCTTGATACAACA
AATCCCGAGCTTTCTGTAGAAACACGACTTCGATTCAAACACAATGTGCAACCTGGA
GTCTTTTATGGTCCATCCACGGATGCTATGGCTCTGTTGAGTAGTAGTAACCCGAAAGAA
GGGTTGCTGCTAGGCCAACAGAAGTTACAGAGTGGTGGATTATGTTGTCAGATGCTGGT
TCCTTAGATGATATAGTCAACTCCACGATGAAGCAGGTGTGA

>G1037 Amino Acid Sequence (domain in AA coordinates: 11-134, 200-248)

MTVEQNLEALDQFPVGMRLVAVDDDTCLKILESLLRHCCYHVTTTNAQKALELLRENK
NKFDLVISDVMDPMDGFKLLELVGLEMDLPVIMLSAHSDPKYVMKGVTGACDYLLKPV
RIEELKNIWQHVVRSRFDKNGSNNGDKRDGSGNEGVGNSDPNNGKGNRKRKDQYNEDE
DEDRDNDNDSQAQKQKRVVWTVELHKKFVAAVNQLGYEKAMPKKILDLMNVEKLTRENV

SHLQKFRLLYLKRISGVANQQAIMANSELHFMQMNGLDGFHHRPIPVGSGQYHGGAPAMRS
FPPNGILGRLNSSSGIGVRSLSPPAGMFLQNQTDIGKFHVSSLPLNHSDDGNNILQGLP
MPLEFDQLQTNNNKSRNMNSNKS IAGTSMAPFSFSTQNSLISAPNNNVVLEGHPQATP
PGFPGHQINKRLEHWSNAVSSSTHPPPAHNSNSINHQFDVSLPHSRPDPLEWNNVSSS
YSIPFCD SANTLSSPALDTTNPRAFCRNTDFDSNTNVQPGVFGPSTDAMALLSSSNPKE
GFVVGGQQLQSGGFMVADAGSLDDIVNSTMKQV*

>G2065 (33..1124)

AACCACACAAAACAAAACAAAACATATTGATGGGGATGAAGAAGGTAAAGCTATCTT
TGATAGCTAATGAAAGATCAAGGAAAACATCCTTCATGAAGAGGAAAAACGGGATATTCA
AGAACTCCACGAGTTGTCAACTCTATGTGGTGTCCAAGCTTGTGCTCTCATCTATAGTC
CATTCAATACCGGTTCCAGAGTCATGGCCGTCAAGGAAGGTGCTAAAAAGGTAGCTTCAA
AGTTTCTGGAGATGCCGCGGACAGCCGAACAGGAAGATGATGGATCAAGAAACCCATC
TTATGGAGAGGATTACCAAAGCAAAAGAGCAACTAAAGAATTGGCTGCTGAGAACCGAG
AATTACAGGTTAGACGATTTATGTTTGATTGTGTTGAAGGCAAAATGTCCCAGTATCGTT
ATGATGCAAAAAGACCTTCAAGATTGCTATCTTGTATGAATCTATATCTCGATCAGCTTA
ACGGAAGGATCGAGTCCATTAAAGAAAACGGTGAGTCGTTGTTGCTTCCGCTCTCTCCTT
TTCCTACTAGAATTGGTGTGACGAAATTGGTGATGAGTCGTTTCCGACTCTCCTATTC
ATTCTACAACCTAGGTTGTAGATACTCCTAATGCTACCAATCCTCATGTTCTTGCGGGCG
ATATGACTCCTTTCTTGATGCGGACGCAATGCGGTAACGTGCTCCAGTCGATTTTCTG
ATCATATTCAATATGAAATATGAATATGAGTCAAAATCTGCATGAACCGTTTCAACACC
TTGTTCTCTACTAACGTTTGTGATTTTATCAAAATCAGAATATGAATCAGGTTCAATACC
AGGCTCCTAATAATCTGTTTAAATCAGATTCAACGAGAATTCTACAACATAAATTTGAATC
TGAATTTGAATCTGAATTCAAATCAGTATCTGAATCAACAACAATCATTATGAATCCGA
TGGTGGAAACAACATATGAATCATGTTGGAGGGCGTGAAAGCATTCTTTCGTGGACAGAA
ACTACTACAACCTAATCAACTACCAGCCGTTGATCTTGCTTCCACCAGTTACATGCCTT
CAACCACCGATGTTTATGATCCTTACATCAACAACAATCTCTAATCACAAAAGACGGAGA
TTTTCTAGTTTAA

>G2065 Amino Acid Sequence (domain in AA coordinates: TBD)

MGMKKVKLSLIANERSRKTSFMKRKNGIFKKLHBLSTLCGVQACALIYSPFIPVPESWPS
REGAKKVASKFLEMPRTARTRKMDQETHLMERITKAKEQLKNLAAENRELQVRRFMFDC
VEGKMSQYRYDAKDLQDLLSCMNLVLDQLNGRIESIKENGESLLSSVSPFPTRIGVDEIG
DESFSDSPIHSTTRVVDTPNATNPVLAGDMTPFLDADANAVTAPSRFSDHIQYENMMS
QNLHEFPQHLVPTNVCDFYQONQNMNVQYQAPNNLFNQIQREFYNINLNLNLNLSNQYL
NQQQSFMNPMVEQHMNHVGGRESIPFVDRNYNYNQLPAVDLASTSYMPSTTDVYDPYIN
NNL*

>G2137 (77..1123)

GGGATTTGACTTTAGCACTTCAAAATCCAAAGCTAAAAGACAAAAAGAATAGAGGTTTCG
ATTTGCATCTCCATTAATGGGCATCGATCTTTCTCTTAAGCTCGAGGCCGAGGAGAAAA
GAAAGAGATAGAAGGATCGAAACATAGCCGTGAGAACAAAGAAGACGAAGAACATGATGC
TAGTGGTGATGAAGATGAACAAATGGTGAAAGAAGACGAAGATGATTCTTCTTTTAGG
TTTAAGAACCCGAGAAGAAGAAAACGAACGTGAAGAGCTCTTGACGTACAGATCCAGAT
GGAAAGTGTAAGAAGAGAATACTAGGTTGAGGAAGCTTGTGAGCAGACTCTTGAAGA
TTATCGTCATCTTGAGATGAAATTCCTCGTTATCGATAAAACCAAGAAGATGGATCTTGA
AATGTTCTCTGGAGTACAAGGCAACGATGTGTGATATAACAAGTAAGGCTCGGAAAAG
AGGAGCTGAGAGATCTCCGTCAATGGAAAGAGAAATAGGGCTTTCACTTTCTTAGAGAA
AAAACAGAAACAAGAAGAGAGCAAGAAGCTGTTCACTCTCATCACCAGATACAATAG
TAGCAGCTTAGATATGAATATGCCACGTATCATTTCATCTTCTCAAGGTAATAGAAAGGC
CAGGGTGTCCGTGAAGGCGAGATGTGAGACCGCAACATGAATGATGGATGCCAATGGAG
GAAGTACGGTCAGAAAACCGGAAAGGGAATCCATGTCTCGAGCTTATTACCGATGCAC
CGTGGCTCCAGGATGTCCCGTTAGAAAACAGGTGCAAGGTGTTTAGAAGACATGTCAAT
ACTGATAACAACCTACGAAGGAACACATAACCATCCACTTCCGGTCGGAGCAACAGCCAT
GGCTTCCACTGCCTCTACTTCTCCATTCTTGTACTCGATTCCAGTGACAACCTCTCTCA
TCCTTCTTATTACAAACTCCTCAAGCCATAGACTCTTCTTTGATTACATACCCACAAAA
TAGCAGCTACAACATCGAACCATAAGAAGCTTGAACCTTTGATGGTCCATCTAGAGGAGA
TCACGTTTCTATCTTCTCAAAACCGATTAAATGGATGATGTAGAGTTTCTATATCTCTA
TGCTTGTCTTTGGTCCCATTATTGTCTATTATGGATTCTTGCCTTTCTTCTGTTCTC
GTTTCTAACATTTATGTTTCGTATA

>G2137 Amino Acid Sequence (conserved domain in AA coordinates:109-168)

MGIDLSLKLEAEKKKEIEGSKHSRENKEDEEHASGDEDEQMVKEDEDDSSSLGLRTRE
EENEREELLQLIQMESVKEENTRLRLKLEQTLQEDYRHEMKFPVIDKTKMDLEMFLGV
QKRCVDITSKARKRGAERSPSMEREIGLSLSLEKKQKQEESSKEAVQSHHQRYNSSSLDM
NMPRIISSSQGNRKARVSVRARCETATMNDGQCWRKYQKTKAGNCPRAYRCTVAPGC
PVRKQVQRCLEMSILITTYEGTHNHPLPVGATAMASTASTSPFLLLDSSDNLSHPSYYQ
TPQAIDSSLITYPQNSSYNRTIRSLNFDGSPSRGDHVSQQNRLNWMW*

>G746 (1..1311)

ATGGGTGAGGAGTTAGCTGACACAATGAACCTGGATTTGAATCTTGGGCCTGGTCCTGAG
TCTGATCTCCAACCTGCACCAACGAGACTGTGAATTTGGCTGATTGGACTAATGACCCG
CCTGAGAGATCTTCTGAAGCTGTGACAAGGATCAGGACTCGGCATAGGACACGGTTCAGA
CAGCTTAATCTCCCGATCCCGGTTCTATCTGAAACCCATACCATGGCTATAGAGCTCAAC
CAGTTGATGGGAAATTCTGTAAATAGAGCTGCTATGCAGACTGGTGAGGGTAGTGAAAGA
GGCAATGAGGATTTGAAATGTGTGAGAATGGCGATGGAGCCCTTGGGGACGGTGTATTG
GATAAGAAAGCGGATGTCGAGAAAAGCAGTGGCAGCGACGGTAACCTTTTCGATTGTAAAT
ATATGTTTGGATTTGTCGAAGGAGCCGGTTCTCACCTGTTGTGGTCATCTTTACTGTTGG
CCTTGCTGTACCAATGGTTACAAATTTCCGATGCAAAGGAATGTCCTGTTGTAAAGGA
GAGGTGACCTCCAAAACCGTGACACCGATCTATGGACGTGGAAACCACAAGAGAGAAATT
GAAGAGAGTTTATGATCTAAGGTCCCATGAGACCACACGCGAGACGCTTGTAGAGCTTG
AGGAATACAAATCAAGGTGCGCTTTTACAATACCAATGGAAGAAATGATTAGACGTATA
CAGAATAGGTTTGACAGGGATTCAACCCAGTCCCTGATTTTAGTAACCGAGAGGCATCA
GAAAGAGTCAACGATCGAGCAATTCGATCCTTAACCGGTTGATGACATCTAGGGGAGTT
AGATCAGAGCAGAACCAGGCTAGTGCTGCAGCAGCAGCCATTGTGCGCAGCATCAGAGGAT
ATTGATCTAAATCCAAACATTGCTCCTGATCTTGAAGGAGAAAGCAACACGAGATTCCAT
CCTCTCTTGATCAGGAGACAGTTACAGTCGCACCGAGTTGCAAGGATCTCGACTTCACT
TCTCGGTTGAGTTTCACTGAGAGGCTTGTGGATGCGTATTTTAGGACTCATCCGTTGGGG
AGGAACCAACAGAGCAAAACCATCATGCTCCTGTTGTGGTTGATGATAGAGACTCATTC
TCAAGCATTGCAGCTGTTATAAACTCTGAGAGTCAAGTGATACTGCAGTTGAGATCGAT
TCTATGGCTCTTTTCGACATCGTCTCGAGGAGAAGGAATGAGAATGGTTCGAGGGTTTCT
GATGTAGACAGTGCAGATTCTCGTCCGCTAGGAGAAGGAGATTTACTTGA

>G746 Amino Acid Sequence (domain in AA coordinates: 139-178)

MGEELADTMNLDLNLGPGPESDLQAPNETVNLADWTNDPPERSSEAVTRIRTRHRTFR
QLNLPPIPVLSETHMTAELNQLMGNVSVNRAAMQTGEGSERGNEDLKMCENGDGLDGV
DKKADVEKSSGSDGNFFDCNICLDLSKEPVLTCGGHLYCWPCLYQWLQISDAKECPVCKG
EVTSTKVTPIYGRGNHKKREIEESLDTKVPMRPHARRIESLRNTIQRSPFTIPMEEMIRRI
QNRFRDSTPVPDFSNREASERVNDRAANSILNRLMTSRGVRSEQNQASAAAAAIVAASED
IDLNPNIAPDLEGESNTRFHPLLIIRQLQSHRVARISTFTSALSSAERLVDAYFRTHPLG
RNHQEQNHHPVVDVDRDSFSSIAAVINSESQVDTAVEIDSMALSTSSRRRNENGSRV
DVDSADSRPPRRRRFT*

>G2701 (46..837)

GTGTTTGTAGTTGAAACTTATTCTTCCCTTTTTTTGTTTGTAGGTATGGAGACTCTGCAT
CCATTCTCTCACCTACCTATCTCTGACCACCGGTTTCGTTGTTCAAGAGATGGTGAGCTTA
CACAGCTCGAGTAGCGGTAGCTGGACTAAAGAAGAGAACAGATGTTTCAACGAGCTCTT
GCGATATACGCTGAAGACTCGCCTGATCGCTGGTTTAAAGTTGCTTCCATGATCCCTGGA
AAGACTGTTTGTATGTTATGAAGCAATATAGTAAGCTTGAAGAAGACGTTTTTCGATATT
GAAGCAGGACGTGTTCCCATTCCTGGTTATCCTGCAGCTTCTTCTCCCTTGGGGTTTGAC
ACGGACATGTGTCGTAAACGGCCTAGTGGAGCTAGAGGATCTGATCAAGATCGAAAGAAA
GGAGTCCCTTGGACAGAGGAAGAACACAGGAGATTCTTGTAGGCCCTTCTCAAGTACGGT
AAAGGAGATTGGAGAAACATATCGAGAAACTTCGTGGTGTCAAAGACGCCAACGCAAGTG
GCGAGCCACGCCAAAAGTATTACCAGAGACAGCTCTCCGGAGCCAAGGACAAACGCAAGG
CCAAGTATCCATGACATCAACCGGCAATCTTCTCAATGCCAATCTCAACCGTTCCCTTT
TCCGATCATAGAGATATTCTCCCTGATTTAGGGTTTATCGATAAGGATGATACGGAGGAG
GGAGTAATAATTATGGGTCAGAATCTCTCTCAGAAAATCTGTTTTCTCCATCACCAACT
TCATTTCGAAGCTGCAATTAACTTCGCCGGAGAAAATGTCCTTCAGTGCCCGGAGCTTAAGGC
AACATAGAATCCCCAACTCAGCGGC

>G2701 Amino Acid Sequence (domain in AA coordinates: 33-81, 129-183)

METLHPFSLPISDHRFVVQEMVSLHSSSSGSWTKEENKMFERLAIYAEDSPDRWFKVA

SMIPGKTVFDMVKQYSKLEEDVFDIEAGRVPIPGYPAASSPLGFDTDMCRKRPSGARGSD
QDRKKGVPTWEEHRRFLLGLLKYGKGDWRNISRNFVVSKTPTQVASHAQKYYQRQLSGA
KDKRRPSIHDIITGNLLNANLNRFSFSDHRDILPDLGFDKDDTEEGVIFMGQNLSSENLF
SPSPTSFEAAINFAGENVFSAGA*

>G1819 (1..639)

ATGGAAGAGAACACGGCAACAACAACCACTACCTGCCGCAACCATCGTCTTCCCAACTG
CCGCCGCCACCATTTGATTATCAATCAATGCCGTTGCCGTCATATTCAGTCCGCTGCCG
TACTCACCGCAGATGCGGAATTATTGGATTGCGCAGATGGGAAACGCAACTGATGTTAAG
CATCATGCGTTTCCACTAACCAAGGATAAAGAAAATCATGAAGTCCAACCCGGAAGTGAAC
ATGGTCACTGCAGAGGCTCCGGTCTTATATCGAAGGCCGTGTGAGATGCTCATTCTTGAT
CTCACAATGCCATCGTGGCTTCATACCGTGGAGGGCGGTGCGCAAACCTCTCAAGAGATCC
GATACGCTCACGAGATCCGATATCTCCGCCGCAACGACTCGTAGTTTCAAATTTACCTTC
CTTGGCGACGTTGTCCCAAGAGACCCTTCCGTCGTTACCGATGATCCCGTGCTACATCCG
GACGGTGAAGTACTTCTCCGGGAACGGTGATAGGATATCCGGTGTGTTGATTGTAATGGT
GTGTACGCGTCACCGCCACAGATGCAGGAGTGGCCGGCGGTGCTGCTGACGAGAGGAG
GCAGCTGGGGAATTGGAGGAAGCAGCGCGGTAATTGA

>G1819 Amino Acid Sequence (domain in AA coordinates: 46-188)

MEENNGNNHLYLPQPPSSQLPPPLYQSMPLPSYSLPLPYSPQMRNYWIAQMGNATDVK
HHAFPLTRIKKIMKSNPEVNMVTAEPVLISKACEMILDLTMRSLHTVEGGRQTLKRS
DTLTRSDISAATRSFKFTFLGDVVPDPSPVTDVPLHPDGEVLPPGTIVIGYPVFDNCN
VYASPPQMGEWPAVPGDGEAAAGEIGSSSGN*

>G1227 (372..1451)

TCTTCCGTGTGTTAACAGAAGTCCCCACAATTGTCTGTCTTCGCTGCGAGACAAAACCTGC
CACAGCCAATAATGTTTCTCTGAGGGACCTTGCTTCTGTCTCAGAGACTCGCTCTCTCTC
CTCTTCTGTCTGTCTCAGCTCTCTCACCAACTCATCTTCAGTCTCAAACAAACATCTG
TTCTCATCTTTGTTTTCTTCTTCTTCTTCTCATATCTCATTTTCAATTTTCCCAATTTT
TCTTCAACATCTTCATAGCAATTTAAGACCACTATTCCATTATAAAGCTAACTGCTTTAG
AAACTCCTCACATTTATTTCTTCCCCATCATTGTTTATAGAGGGAGAAAGAAAAGAGC
TCAGCTTTCTGATGGAGAGGAGTATTCAAGGACAAAACAAGCTCTGTTGTTGGACAAA
AAGTGAATGTGAGAAGAAGCCTACAAGTTCAAGAACTGTAGAGGATCATCAAAGCTTTG
CCCTTGAAGAGGAAGAACAACAACCTCTCAACTCCGAGCTTGCTGCAAGACACAACAATAC
CATTTCTACAAATGCTGCAACAAGTGAAGACCCTTACCCTTTTTGTTCATTCAAAGACC
CAAGCTTTCTAGCACTACTATCTCTCCAGACACTTGAAAAGCCTTGGGAACTCGAAAAC
ACCTCCCACATGAAGTTCCAGAGTTTCATTACCGATCCATTCTGAAACCAACCACTACT
ATCATAATCCATCTTTGGGAAGGAGTCAATGAAGCCATCTCAAACCAAGAACTTCCATTCA
ACCCACTAGAGAATGCGCGTTCAAGACGCAAGCGGAAAAACAACAACCTTGGCATCATTGA
TGACAAGAGAAAAGCGAAAGAGAAGAAGAACTAAACCAACAAGAACATAGAAGAGATAG
AGAGTCAAAGAATGACACACATTGCGGTTGAACGAAACCGCAGACGCCAAATGAACGTTT
ATCTGAACCTCACTCCGCTCCATCATTCCATCTTCATACATCCAGAGGGGAGACCAAGCGT
CAATAGTAGGAGGAGCAATAGACTTCGTAAAGATCCTAGAGCAACAGTTGCAATCCCTTG
AAGCACAAAAGAGAAGTCAACAGAGTGATGATAACAAGAGCAAAATCCAGAAGATAACA
GTCTCAGGAACATTTTCGTGCAACAAGTTGCGTGCGAGTAATAAAGAAGAACAAAGTAGCA
AACTCAAATCGAAGCCACAGTGATAGAGAGTCACGTCAACCTAAAAATTCATGTACGA
GGAAACAAGGACAACCTTCTCAGATCAATCATATTGCTGGAGAACTTCGATTCACTGTTT
TTCATCTCAACATCACATCTCCGACCAATACATCTGTCTCTTATTCTTCAACCTCAAGA
TGGAAGATGAATGTAATTTGGGATCAGCGGATGAGATAACGGCGGCGATTTCGTGAGATT
TCGACAGCTGATTGACTAATCCAAGTAAAAAGTAAAAATAAAAAAAGAAACGTTTACTTTG
GTAACCTCGTTTTTCATGATTAAATCTTTATTTGGTCTGATGTGATTGGAGTCTTCTCGG
CATGGAACCTTGACTTTTGGTTTTAGGGTACTAGTCTCTACAGAAGCTGTGGTCTTCTTTG
GATGC

>G1227 Amino Acid Sequence (domain in AA coordinates: 183-244)

MERSIQGNKLCCLDQKVNRRSLQVQETVEDHQSFALBEEEQQLSTPSLLQDTTIPFLQ
MLQQSEDPSPFLSFKDPSFLALLSLQTLKPEWELNYLPHEVPEFHSPHSETNHYYHNP
SLEGVNEAISNQELPFNPLENARSRRKRKNNNLASLMTREKRKRRTKPTKNIEEIESQR
MTHIAVERNRQRQMNVLNLSLRSIIPSSYIQRGDQASIVGGAIDFVKILEQQQLQSLEAQK
RSQSDDNKEQIPEDNSLRNISSNKLRSNKKEQSSKLKIEATVIESHVNLKIQCTRKQG
QLRSIILLEKLRFVTLHLNITSPTNTSVSYFNLKMEDECNLGSADEITAAIRQIFDS*

>G2417 (118..1311)

CATACCGGTGGAAGATTCTGCTTTACTACGCTCTCCGCTTCTTCTCTCTCGATTCGAT
TCTCCTCATGGGTTTATCATGAATTTTAGGTTTTGAGTAATTCAGAACTCGAGTGATG
ATCCCGAATGATGATGATGATGCAAATTTCTATGAAGAATTATCCGTTAAATGATGATGAT
GCAAATTTCTATGAAGAATTATCCGTTAAATGATGATGATGCAAATTTCTATGGAGAATTAT
CCGTTAAGGTCAATTCGACGGAGCTTTACACACTTGTTCATTGATACCACCTTCTTTA
CCAAACCTTCAGAAAGCAGCAGCAGACATGTCCTTCAATTCAGAACTCAATCAAATCATG
GCAAGGCCTTGTGATATGCTCCCTGCCAATGGTGGAGCTGTTGGTCATAACCCTTTTGTG
GAACCAGGATTCAACTGCCCGGAGACAACAGATTGGATTCCCTCTCCACTCCCCATATT
TATTTTCTTCGGGTTCTCCCAATCTAATAATGGAGGATGGTGTCAATTGATGAGATTCAC
AAACAAAGTGACTTGGCACTTTGGTATGACGACTTGATTACCACTGATGAAGATCCACTC
ATGTCTAGTATCTTGGGCGATCTTCTCTTGACACTAATTTCAACTCAGCTTCAAAGGTC
CAGCAACCAAGTATGCAATCGCAGATTCAACAACCCCAAGCTGTTCTGCAGCAGCCTTCT
TCTTGTGTGGAATTGCGCCCACTTGATAGGACAGTATCCTCAAACAGCAACAACAATAGC
AACAGTAATAATGCAGCAGCAGCAGCTAAGGGACGTATGCGTTGGACGCCTGAACTTCAT
GAGGTTTTTGTGACGCTGTTAACCAGCTCGGTGGCAGTAATGAAGCAACTCCTAAAGGT
GTCTTGAAGCATATGAAAGTCGAAGGTTTGAATTTTTTCATGTCAAAGTCATTTGCAG
AAATATAGAACAGCTAAATATATATACCAGTACCATCAGAAGGTTGCGCGGAGGCAAGGTTG
ACACCGCTTGAGCAAATTACATCTGATGATACGAAACGTGGGATAGATATCACTGAGACT
CTGCGAATTCAGATGGAACATCAGAAGAACTGCATGAGCAGCTTGAGAGTCTAAGAACA
ATGCAACTTCGGATAGAAGAGCAAGGAAAGGCGCTGTGATGATGATTGAGAAGCAAAAT
ATGGGTTTTCGGCGGACCAGAACAAGGAGAGAAAACAGTGCGAAAACGCCTGAAAATGGT
TCAGAGGAGTCGGAATCCCCGCGGCCAAAGCGTCCGAGAAATGAAGAATGAAGGAAACCT
TTCTTCGATGGTAGATCATAAACTGTGGTTTTTGGTGGAGTTGTAGAGTATGACTTATT
AGGAGTAGAGCTTTTCAGTCTTCTTCAGGC

>G2417 Amino Acid Sequence (domain in AA coordinates: 235-285)

MIPNDDDDANSMKNYPLNDDDDANSMKNYPLNDDDDANSMENYPLRSIPTELSHTCSLIPPS
LPNPSEAAADMSFNSELNQIMARPCDMLPANGGAVGHNPFLPEGFNCPETTDWIPSLPH
IYFPSGSPNLIMEDGVDEIHKQSDLEPLWYDDLITDEDPLMSSILGDLILLDTNFNSASK
VQQPSMQSQIQPPQAVLQQPSSCVLRPLDRTVSSNSNNNSNNNAAAAAKGRMRWTPPEL
HEVFVDVAVNLGGSNEATPKGVLKHMKVEGLTIFHVKSHLQKYRTAKYIPVPSEGSPEAR
LTPLEQITSDDTKRGIDITETLRIQMEHQKKLHEQLSLRTMQLRIBEQGKALLMMIEKQ
NMGFGGPEQGEKTSAKTPENGSESESPRPKRPRNEE*

>G2116 (104..1117)

TTTCATCTCCATCATATCTCCATTGACATTGTTCTCAATTGCGAATAATAATCATAATTA
TTCACACAACCAAAGCATTCATCTCTCAGATTCTCTTAAAAAATGGAGAAATCAGATCC
TCCACCAGTCCCAAAGCCCGCGCCACTATTATCCCTCTCTCCGATCCAATTCCTAATGC
CGATCCGATTCCATCTCTCTCTTCCACCGCGATCTCGCTCCGACGATATGTCCATGTT
CATGTTTCATGGATCCCTCTCTCTCCCGCCGACCACTTCTCTCCGACGACCTTCCCTCCGA
CGACGATCTCTCTCTCTCTTTCATCGATGTGATAGCCTCACCTCTAATCCCAATCCCTT
TCAAAATCCTTCCCTCTCTCTCAACTCCGTTTCCGGCGCTGCTAATCTCTCTCTCTCTC
TTCTCTCTGCTCTCTCGCCACCGTCACAGCAATTCGTTGACGCTGGATGCGCCATGTATGC
CGGTGATATCATGGACGCTAAGAAAGCTATGCCTCTGAAAACTCTCTGAGCTTTGGAA
CATCGATCCCAAACGCGCCAAAAGGATTCTAGCGAATCGACAATCTGCAGCTCGATCCAA
AGAGAGAAAAGCTCGATACATTCAAGAACTTGAGCGCAAAGTTCAATCTCTTCAAACCGA
AGCTACCACTCTCTCTGCTCAGCTTACTCTCTACCAGAGAGACACAAATGGACTAGCAAA
CGAAAAACACAGAGCTGAAACTTAGGTTGCAAGCAATGGAACAACAAGCTCAGCTTCGTAA
TGCTTTAAACGAAGGTTGAGGAAAGAAGTTGAAAGGATGAAGATGGAGACAGGAGAAAT
CTCTGGTAATTCAGATTCTGTTTGTATGGGAATGCAGCAGATTCAATCTCTCTCTCAAC
TTTCATGGCTATTCACCATATCATGGCTCAATGAACCTCCATGATATGCAGATGCATTCT
TAGTTTCAATCTATGGAGATGTCGAATCTCAAAGCGTGTGCGACTTTCTACAGAACGG
CCGAATGCAAGGCTGGAGATTAGTAGCAATAGCTCAAGCTTAGTCAAATCTGAAGGACC
TTCTCTCTCTGCTAGTAGAGTAGCTCTGCCTATTGACGACAAGATTATGATGAGGCTCA
TTTTTCTG

>G2116 Amino Acid Sequence (conserved domain in AA coordinates:150-210)

MEKSDPPPVPKPGATIIPSSDPIPNADPIPSSSFHRRSRSDMSMFMDPLSSAAPPSS
DDLPSDDDLFSSFDVDSLTSNPNPFQNPSSLSSNSVSGAANPPPPSSRPFRHSNSVDA

GCAMYAGDIMDAKKAMPPEKLSELWNIDPKRAKRILANRQSAARSKERKARYIQELERKV
QSLQTEATTLAQLTLYQRDTNGLANENTELKLRLQAMEQQAQLRNALNEALRKEVERMK
METGEISGNSDSFDMGMOQIQYSSSTFMAIPYPYHGSMNLHDMQMHSSFNPMEMSNSQSVS
DFLQNGRMQGLEISSNSSSLVKSEGPSLSASESSAY*

>G647 (1..948)

ATGATGATCGGCGAAAAATAAAACCGGCCACATCCAACGATCCATATCCCTCAATGGGAT
CAAATCAACGATCCAACCGGCCACATCTCTTACCATTCTCTTCCGTCAACCTTAACAGC
GTTAACGACTACCCACACTCTCCGTCACCGTATCTCGACTCCTTCGCTTCTCTTCCGT
TACCTCCCGTCAAACGAGTTAACAAACGATTTCAGACTCATCAAGTGGCGACGAGTCATCA
CCACTACCGACTCATTCTCTCCGACGAGTTTCGCATCTACGAGTTCAAAATCCGGCGA
TGCGCTCGAGGTCGATCTCATGATTGGACGGAGTGTCCGTTTCGCACATCCCGGAGAAAAA
GCTCGACGACGTGATCCGAGAAAGTTTCTTACTCCGGCACCGCTTGTCTCGAGTTTCGT
AAAGGAAGTTGTAGAAGAGGTGATTTCGTGTGAGTTCTCTCATGGAGTTTTCGAGTGTGG
CTCCATCTTCTCGTTACCGTACTCAGCCGTGTAAAGACGGAAGTAGCTGCCGAGAGAAGA
ATCTGTTTCTTCGCTCATACGACGAGCAGTTACGTGTATTACCTTGTTCGTTAGATCCA
GATCTTGGATTCTTCTCAGGATTAGCTACTTCTCCGACTTCGATTCTTGTTCCTCTTCG
TTTTACCACCGTCGGAATCTCCGCCGCTTCTCCGAGTACCGGTGAACCTATTGCGTTCG
ATGAGGAAAAATGCAATTGAACGGAGGTGGTTGTTCGTGGAGTTCTCCGATGAGATCTGCA
GTTAGGTTACCTTTTTCGTCTCTTCGCTCCGATTTCAGGCGGCAACGTGGCCGAGGATA
AGAGAGTTTGAGATCGAAGAAGCTCCGGCGATGGAATTTGTGGAATCTGGGAAAGAGCTG
AGAGCGGAGATGTATGCAAGACTCAGTAGAGAGAACTCACTCGGTTGA

>G647 Amino Acid Sequence (domain in aa coordinates: 77-192)

MMIGENKNRPHTIHIPQWDQINDPTATISSPFSSVNLNSVNDYPHSPSPYLDSEFASLFR
YLPSNELTNDSDSSSGDESSPLTDSFSSDEFRIYEFKIRRCARGRSHDWTECPFAHPGEK
ARRRDRPKFHYSGTACPEFRKGSCRRGDSCEFSGVFECWLHPSRYRTQPKDGTSCRRR
ICFFAHTTEQLRVLPCSLDPDLGFFSGLATSPTSILVSPSFPSPSESPPLSPSTGELIAS
MRKMQLNGGGCSWSSPMRSVRLPFSSSLRPIQAATWPRIREFEIEEAPAMEFVESGKEL
RAEMYARLSRENSLG*

>G974 (377..1162)

AAAAAAAAGTTGATATACTTTCTGGTTTTCTCCTTAACTTTTATTCTTTACAAATCCAT
CCCCCTTAGATCTGTTTATTTCCCGCTACTTTGATTCAATTTCTGTTAGTAATCTGTCTTT
CGTATAGAAGAAAACGATTTCTTGGTTTGTATTTCTTAAAGAGATCAATCTTTTTTTA
TTTTTGATCTTCTTGTGTTTTTTTCTTTGTAGAATTAATCGTTTGTGAGGGTATTTTT
TTAATTCCTCCTCTCAGAAATCTACACAGAGGTTTTTTATTTTATAAACCTCTTTTTTCG
ATTTTCTTGAAAACAAAAAATCCTGTTCTTTACTTTTTTACAAGAACAAAGGAAAAAAA
TTTCTTTTTATTAGAAATGACAACCTTCTATGGATTTTTACAGTAACAAAACGTTTCAACA
ATCTGATCCATTTCGGTGGTGAATTAATGGAAGCGCTTTTACCTTTTATCAAAAGCCCTTC
CAACGATTCATCCGCGTTTGCCTTCTCTACCCGCTCCAATTTATACGGGTTCGGATCT
CCACTCATTTTCTCACCATCTTAGTCCTAAACCGGTCTCAATGAAACAAACCGGTACTTC
CGCGGCTAAACCGACGAAGCTATACAGAGGAGTGAGACAACGTCACTGGGGAAAAATGGGT
GGCTGAGATTCTGTTTACCGAGGAATCGAATCGACTTTGGCTCGGAACATTCGACACGGC
GGAGGAAGCTGCTTTAGCTTATGACAAGGCGGCGTATAAGCTCCGAGGAGATTTGCGCG
GCTTAATTTCCCTGATCTCCGTCATAACGACGAGTATCAACCTCTTCAATCATCAGTCGA
CGCTAAGCTTGAAGCTATTTGTCAAACTTAGCTGAGACGACGCAGAAACAGGTGAGATC
AACGAAGAAGTCTTCTCTCGGAAACGTTTCATCAACCGTCGCACTGAAACTACCGGAGGA
GGACTACTCTAGCGCCGGATCTTCGCCGCTGTTAACGGAGAGTTATGGATCTGGTGGATC
TTCTTCGCCGTTGTTCGGAGCTGACGTTTGGTGATACGGAGGAGGAGATTACGCCGCCGTG
GAACGAGAACGCGTTCGAGAAGTATCCGTCGTACGAGATCGATTGGGATTCGATTCTTCA
GTGTTTCGAGTCTTGTAAATTAGATGTTGCCATAGGGGTATTTTAGGGACTTTAGAGCTCT
CTGCGATGGAGTTTTTGGTCATTGCAGAGATTTTATTATTATTAAGGGGGTTTGTATGT
TAATATCAATAAGTTTATCTACTTTGATGTTAATTAGTGTTAATCTCTGCGTCGGTCCA
AGCTGTTTTTTTTTGGCATGCTTCGACCGTGTGAGATTTCTTATGTAATTTTTGTAGTTC
CTTGATTTTCTTAGTTCAAGTTAAATTGGCACAAAAA

>G974 Amino Acid Sequence (domain in AA coordinates: 81-140)

MTTSMDFYSNKTFFQSDPFGGELMEALLPFIKSPSNDSSAFAPSLPAPISYGSDDLHSFSH
HLSPKPVSMKQGTGSAAKPTKLYRGVQRHWGKWVAEIRLPNRTRLWLGTDFDTAEAAAL
AYDKAAYKLRGDFARLNFDPDLRHNDYQPLQSSVDAKLEAICQNLAEATTQKQVRSTKSS

SRKRSSTVAVKLPEEDYSSAGSSPLLTESYSGSGSSSPLSELTFGDTEEEIQPPWENENAL
EKYPSYEDWDSILQCSSLVN*

>G1419 (27..692)

GAAGACTCCAACATAATTCATCATCTATGGCTTCTTCACATCAACAACAGCAAGAACAAG
ACCAGTCAGCTTTAGATCTCATAACCCAACACCTTCTTACTGATTTCCCTTCCTTAGACA
CCTTTGCCTCCACCATCCACCCTGCACCACCTCAACTCTAAGCCAACGCAAACCCACCTC
TTGCCACTATAGCAGTTCCTACTACTGCACCGGTGGTTCAAGAGAATGATCAAAGGCATT
ACAGAGGCGTCAGGAGAAGACCATGGGGTAAGTATGCGGCTGAGATCAGAGACCCAAACA
AGAAAGGTGTTCTGTCTGGTTAGGCACCTTTTGACACAGCCATGGAAGCTGCAAGAGGTT
ATGACAAGGCAGCTTTTAAACTACGAGGAAGCAAAGCTATTCTTAACTTCCCACTTGAAG
CAGGAAAGCATGAGGACTTGGGAGACAACAAGAAGACTATTTCTTTAAAAGCAAAGAGGA
AGAGACAGGTGACGGAGGATGAAAGCCAGCTGATCAGCCGTAAAGCTGTTAAGAGGGAAG
AAGCTCAGGTTGAGGCTGATGCTTGTCCATTACGCCATCAAGTTGGAAGGGGTTTGGG
ACGGAGCAGACAGTAAAGACATGGGAATATTTCCGTGCCTCTGTTATCTCCTTGTCCAT
CTCTTGGACACTCTCAACTCGTAGTTACTTAAGCTTCAGAGGGTCAAAGTGGAAAAATC
AACATTGGATTGTTTTCAAAGCTTCTAGATTAGCTGATTGTAAAAAATGTTTTACTATA
TTCATTCTTCTTAAATGCAATCTTTCTACCTTCC

>G1419 Amino Acid Sequence (domain in AA coordinates: 69-137)
MASSHQQQEQDQSDLDLITQHLLTDFPSLDTFASTIHHCTSTLSQRKPPLATIAVPTT
APVVQENDQRHYRGVRRRPWGKYAAEIRDPNKKGVRVWLGTFTDAMEAARGYDKAAFCLR
GSKAILNFPLEAGKHEDLDGNKKTISLKAKRKRQVTEDESQILSRKAVKREBAQVQADAC
PLTPSSWKGFWDGADSKDMGIFSVPLLSPCPSLGHSQLVVT*

>G1634 (22..855)

TTATCTCGTAGCCTTTAAACGATGGAGACTCTGCATCCACTACTCTCGCACGTGCCAACT
TCTGACCACCGGTTGTAGTTCAAGAGATGATGTGCTTGCAAAGCTCGAGCTGGACTAAA
GAAGAGAACAAGAAGTTTGAGCGAGCTCTTGCTGTCTACGCTGATGACACGCTGATCGC
TGTTTCAAAGTTGCTGCTATGATCCCTGGAAGACCATATCAGATGTCATGAGGCAATAC
TCTAAGCTTGAAGAAGACCTCTTCGATATCGAAGCAGGACTTGTCCCGATCCCGGGTTAC
CGTTCAGTTACTCCTTGTGGATTGATCAGGTTGTGAGTCCACGTGACTTTGATGCGTAT
CGTAAACTTCTTAATGGAGCCAGAGGATTGATCAAGACCGTAGGAAAGGAGTTCCATGG
ACGGAGGAAGAACAACAGGAACTTTGTGGGATCAAAAACACCAACTCAGGTTGCAAGTCATGCC
CAAAAGTACTACCAAAGACAGCTTTCCGGTGCGAAAGACAAACGACGGCTAGCATTAC
GACATCACCACCGTCAATCTTCTCAATGCCAATCTTAGCCGTCCATCGTCTGATCACGGT
TGCTTAGTCTCAAAACAGGCCGAGCCGAAACTAGGGTTCACCGACAGGGATAATGCAGAG
GAGGGAGTTATGTTTCTTGGTCAGAATCTATCCTCGGTCTTCTCTTCTACGATCCTGCC
ATTAAGTTTTCCGGAGCAAATGTTTACGGTGAAGGAGTTACTGTATCTCACAAGATCTT
GAAACGAGAAAATGAAATTTTGAATTTTAACTATTGCAACGAAACCATAATTGC

>G1634 Amino Acid Sequence (domain in AA coordinates: 129-180)
METLHPLLSSHVPTSDHRFVVQEMMCLQSSWTKEENKFERALAVYADDPDRWFKVAAM
IPGKTISDVMRQYSKLEEDLFDIEAGLVPIPGYRSVTPCGFDQVVS PRDFDAYRKLPGA
RGFDQDRRKGVFWTEEEHRRFLLGLLKYKGDWNRNISRNFVGSKTPTQVASHAQKYQRQ
LSGAKDKRPSIHDITTVNLLNANLSRPSDHDGCLVSKQAEPKLGFTDRDNAEEGVMFLG
QNLSSVFSSYDPAIKFSGANVYEGGGYICISQDLETRK*

>G1637 (1..954)

ATGGTGAAGGAGACGGTGACGGTGGCGAAAACGTGCTCACACTGTGGCCATAATGGCCAT
AACGCACGGACTTGTCTCAACGGCGTTAATAAGGCAAGTGTTAACTGTTTCGGCGTTAAT
ATATCGTCTGATCCGATTAGGCCGCTGAGGTAACGGCGTTAAGGAAGAGTCTTAGTTTG
GGAAACCTTGATGCTCTTCTCGCTAACGATGAAAGTAACGGTAGCGGTGATCCTATCGCC
GCCGTTGATGATACCGTTATCATTCGGATGGTCAGATTCAATCCAAGAAGGGTAAACT
GCTCATGAGAAGAAAAGGGGAAGCCATGGACGGAAGAAGAACATCGTAATTTCTTAATC
GGTTTAAACAAACTCGGAAAAGGAGATTGGAGAGGCATTGCAAGAGTTTCGTGTCGACA
AGAACACCAACACAAGTCGCAAGTCATGCTCAGAAAATATTTTATTAGGTTAAACGTTAAC
GACAAGAGAAAAGACGTGCTAGTCTCTTTGACATCTCTCTCGAAGATCAGAAGGAGAAA
GAGAGGAATCTCAAGATGCTTCAACAAAGACTCCACCTAAACAACCAATAACCGGAATT
CAACAACCGGTAGTACAAGGTCATACTCAAACCGAGATTTTCAACAGGTTTCAGAATTTA
TCAATGGAGTATATGCCAATCTACCAACCCATACCACCTTACTACAACCTTCCACCTATT

ATGTACCATCCAAATTATCCAATGTACTATGCCAACCTCAAGTACCGGTTAGGTTTGT
CATCCTTCTGGTATACCTGTTTCCAAGACATATACCGATTGGTTGCCTCTGTCTCAACCG
AGTGAAGCTTCTAATATGACAAATAAAGACGGTTTGGATCTTCATATCGGTTTGCCTCCA
CAAGCTACTGGAGCTTCTGACTTGACTGGTCATGGCGTTATTCATGTGAAATGA
>G1637 Amino Acid Sequence (domain in AA coordinates: 109-173)
MVKETVTVAKTCSHCHGNHARTCLNGVNKASVKLFGVNISSDPIRPPEVTALRKSLSL
GNLDALLANDESNGSGDPIAAVDDTGYHSDGQIHSKKGKTAHEKKKGKPWTEEHNRNLI
GLNKLKGDWGRGIKSFVSTRPTQVASHAQKYFIRLNVNDKRKRASLFDISLEDQKEK
ERN SQDASTKTPPKQIPITGIQQPVVQGHQTQTEISNRFQNL S MEYMPIYQPIPPYYPFPPI
MYHPNYPMYANPQVPVRVHPSPGIPVPRHIPIGLPLSQPSEASNMTNKDGLDLHIGLPP
QATGASDLTGHGVIHVK*

>G1818 (601..1161)
TAACAAATCAAATAATTAGAGAAATAACCAAAATTTAACTTTTAGAGGGACTACAGGATT
TGTACTTTGTACATTCATATATTATTGTTATATATCGTTTCATACATTAATTTGAACCAA
TGTAATTAAGTAAATTCATTTTAACATCATGAGCAAATCTTATTAATTAATCTCTTAA
AATTTTGAAGCAAATTAATGCTTTTACATTTTGAACAAATCATTTTGAACAGATA
TTCAAACTAAGTTTGTACAGCAAATTTTAACTTTCAATTTTATAGAGAAAAAGGTAT
TTTTTTTTTGTTCATTTTATAAGACTATTATTGGTATATAATATACACTTTAAGTA
AAAACAAATCTCTTTCTTTTCTTATAATAACCAACCACAAGTCTGTCAGTCACACA
CATACAGTTAATAACATTAATATTTCTTAACAAACTACTAAATAGGTTGAGATTCATATA
TGTAAGAGATCACTTCTTAATCTTATCCTACCATATCTTATATACGCTTAATTTTCTT
TATATATGCAAACCTCCACATAAAAAATATCTCAAACCCAAACACTTCAAACAAAAA
ATGGAGAACAAACAACAACCAACCAACAGCCACCGAAAGATAACGAGCAACTAAAGAGT
TTCTGGTCAAAGGGGATGGAAGGTGACTTGAATGTCAAGAATCACGAGTTCCCATCTCT
CGTATCAAGAGGATAATGAAGTTTGATCCGGATGTGAGTATGATCGCTGCTGAGGCTCCA
AATCTCTTATCTAAGGCTTGTAAGTGTGTCATGGACCTCACGATGCGTTTATGGCTC
CATGCTCAAGAGAGCAACCGCATCACGATACGGAATCTGATGTGATGCCGTAGTGTCT
CAAACCGTCATCTTTGATTTCTTGCGTGATGATGTCCCTAAGGACGAGGGAGAGCCGTT
GTCGCCGCTGCTGATCCTGTGGACGATGTTGCTGATCATGTGGCTGTGCCAGATCTTAAC
AATGAAGAACTGCCGCCGGGAACCGTGATAGGAATCCCGTTTGTACGTTTAGGAATA
CACGCCGCACACCCGAGATGCCGAGGCTTGGACCGAGGAGGATGCGACTGGGGCAAAT
GGAGGAAACGGTGGGAATTAATATTGGATTGGGTTTGTAAACCGCTGTTGTGAGAACTT
GAATTTCTTTTGTAGTTCTGCTTATGTTTCAATGTTATGTTTGTAGTTGTGAATGTA
TTTCTGTTGTTTGTCCAAAAAAGAAATGTTTCTGTTGTTGTTCTTCAAATGA
ATCTAATGGTTTATGAATATTGGCTTTAGATTAAATTTATGCATACAAAAACACAAGGATT
ACGGATAAAAAAGTCTCAGTTTACCCATGGAAACATAATCTTCTAGTGATTCTTATGA
GAGTAGAAAAAGATCATATATTATAATCTATTTTCATAAGAGATAGGGTACTGTAAACAAG
GATGTTTATTCGGCTATTTCTTTTAAATCACTTTTACTTGTCAAGACTCTTTTGT
GTTTGCAGCTTTTGTAGATTACATCTAGAGGCAACAAGATCCAGAGATCTAGCAAAA
AAAACCTTATTTGAAACCTGAATCTATTTTAAAAATTTTCAACTCATTTTTCGTTCTTA
TTCTTTGTTTTCACCGAATTTGGCGCACAAACGATTTATTGAATTTTGTCTTTCAAG
>G1818 Amino Acid Sequence (domain in AA coordinates: 36-113)
MENNNNNHQPPKDNELKSFWSKMEGDLNVKNHEFPISRIKIMKFDPDVSMIAAEAP
NLLSKACEMFVMDLTMRSLHAQESNRLTIRKSDVDAVVSQTVIFDFLRDDVPKDEGEFV
VAAADPVDDVADHVAVPDLNNEELPPGTVIGTPVCYGLIHAPHPQMPGAWTEEDATGAN
GGNGGN*

>G1820 (1..609)
ATGGCTGAGAACAAACAACAACGGCGACAACATGAACAACGACAACCACCAGCAACCA
CCGTCGTACTCGCAGCTGCCGCCGATGGCATCATCAACCCCTCAGTTACGTAATTACTGG
ATTGAGCAGATGGAACCGTCTCGGATTTCAAAACCGTCAGCTTCCATTGGCTCGAATT
AAGAAGATCATGAAGCTGATCCAGATGTGACATGGTCTCCGAGAGGCTCCGATCATC
TTCGCAAAGGCTTGCGAAATGTTTATCGTTGATCTCACGATGCGGTCTGCGCTCAAAGCC
GAGGAGAACAAACGCCACAGCTTCAGAAATCGGATATCTCAACGCAGTGGCTAGCTCT
TTACACTACGATTCTTGTGATGTTGTCCTAAGGACGAGTCTATCGCCACCGCTGAT
CCTGGCTTTGTGGCTGATGCCACATCCTGACGGTGGAGGATACCGCAATATTATTATCCA
CCGGGAGTGGTGTGAGGAACTCCTATGGTTGGTAGTGGAATGTACGCCCATCGCAGGCG
TGGCCAGCAGCGGCTGGTGACGGGGAGGATGATGCTGAGGATAATGGAGGAAACGGCGGC

GGAAATTGA

>G1820 Amino Acid Sequence (domain in AA coordinates: 70-133)
MAENNNNGNDNMNNDNHQQPPSYSQLPPMASSNPQLRNYWIEQMETVSDFKNRQLPLARI
KKIMKADPDVHMVSAEAPIIFAKACEMFIVDLTMRSLWKAENKRHTLQKSDISNAVASS
FTYDFLLDVVPKDESIATADPGFVAMPHPDGGGVPPQYYPGVMGTPMVGSGMYAPSQA
WPAAAGDGEDDAEDNGNGGGN*

>G1903 (1..1200)

ATGTCTAAATCTAGAGATACGGAGATAAAGTTGTTTGGGAGGACAATCACATCTCTTTTA
GATGTGAATTGTTATGATCCGTCGTCGTTGTCCCCTGTTACGATGTTTCTTCTGATCCA
AGCAAGGAGGATTCTGCTTCTTCTTCTTCTGTTCTCCAATATTGGACCAATCAGG
GTTCCGGTTAAAAAAGTGAGCAAGAGAGTAACAAATTCAAAGATCCATATATATTATCC
GATCTAAACGAACCACAAAAGCAGTATCTGAGATTTTCATCACCAAGAAGTTCCAAGAAC
AACTGTGATCAACAGAGCGAGATCACAACAACACTACCACAAGTACTACATCAGGAGAG
AAATCAACGGCTCTCAAGAAACCGGACAAGCTTATTCATGTCTTAGATGTGAAAGCGCA
AACACCAAATTCTGTTATTACAACAACACTACAACGTGAACCAGCCACGTTACTTCTGCAGG
AACTGTGAGAGTATTGGACAGCTGGTGGATCTATGAGGAACGTTCTGTTGGCTCAGGT
CGTCGCAAGAACAAAGGATGGCTTCTTCAAACCATTACTTGCAAGTCACTTCTGAGGAT
TGTGATAATAATAACTCGGGGACGATCCTTAGTTTTCGGTTCTTTCGGAGTCTTTCGGTTACA
GAGACTGGTAAGCATCAGTCAGGTGATACAGCAAAGATAAGTGCTGATTCACTTTCTCAA
GAAATAAAAGCTACCAAGGGTTTCTTCTCCGCAAGTAATGTTACCTAATAATTCTTCT
CCTTGGCCTTACCAATGGAGTCCAACGGGTCTTACGCTAGTTTCTACCCTGTCCCCTTC
TACTGGGGATGCACGGTTCCGATATACCCTACCTCAGAGACTTCATCATGTTTAGGAAAA
CGGTCAAGAGATCAAACTGAAGGAAGAATCAATGATACTAATAACAATAACTACTACA
AGAGCAAGATTGGTCTCAGAATCTCTTAGAATGAATATCGAAGCTAGTAAGAGCGCTGTG
TGGTCTAAGTTACCGACAAAACCCGAGAAAAAACGCAAGGATTCACTTTGTTCAATGGA
TTTGACACAAAGGAAACAGCAACAGAAGTAGCTTGGTCTCCGAAACTTCTCACAGTCTA
CAAGCAAACCTGACGCGATGTCTAGAGCTATGAACCTCAGGGAGAGCATGCAACAATAA
>G1903 Amino Acid Sequence (domain in AA coordinates: 134-180)
MSKSRDTEIKLFGRITITSLLDVNCYDPSSLSPVHVDVSSDPSKEDSSSSSSSSCSPTIGPIR
VPVKKSEQESNKFDPYILSDLNEPPKAVSEISSPRSSKNNCDQQSEITTTTTTSTTSGE
KSTALKKPKDKLIPCRCESANTKFCYNNYNVNPQRYFCRNCQRYWTAGGSMRNPVVGSG
RRKNKGWPPSSNHYLQVTSEDCDNNNSGTILSFGSSESVTETGKHQSGDTAKISADSVSQ
ENKSYQGFLPPQVMLPNNSSWPYPQWSPTGPNASFYPVPFYWGCTVPIYPTSETSSCLGK
RSRDQTEGRINDTNTTITTTTRARLVSESLRMNIEASKSAVWSKLPKPEKKTQGFSLFNG
FDTKGNNSNRSSLVSETSHSLQANPAAMSRAMNFRESMQQ*

>G371 (1..582)

ATGGAGATTGAGAAGGATGAGGACGACACAACATTGGTTGATTCTGGAGGAGACTTCGAC
TGCAACATATGTTTGGATCAGGTTTCGAGACCCGGTCGTGACTTATGTGGCCACCTGTTT
TGTTGGCCCTGCATTACAAGTGGACTTATGCGTCCAACAATTCAAGACAACGAGTCGAT
CAATACGATCATAAGAGGGAACCAACAAAATGTCCGGTATGCAAATCTGATGTCTCCGAG
GCTACGCTTGTCCCGATCTACGGACGAGGACAGAAAGCTCCCAGTCCGGTTCAAATGTA
CCGAGCAGACCAACTGGTCCGGTTTATGACTTAAGAGGAGTTGGTCAACGTTTAGGAGAA
GGGGAGAGTCAACGTTACATGTATAGAATGCCTGATCCGGTGATGGGTGTGGTATGCGAA
ATGGTATACCGGAGACTATTTGGAGAGTCTTCGAGCAACATGGCACCTTACCGCGATATG
AATGTCCGGTCTAGGCGACGGGCAATGCAGGCTGAGGAGTCATTAAGCAGAGTCTACTTG
TTTCTACTTTGCTTTCATGTTTATGTGTCTATTTCTCTTCTAA

>G371 Amino Acid Sequence (domain in aa coordinates: 21-74)
MEIEKDEDDTTLVDSGGDFDCNICLDQVRDPVVTLCGHLFCWPCIHKWTYASNNRQRVD
QYDHKREPPKCPVCKSDVSEATLVPIYGRGQKAPQSGSNVPSRPTGPVYDLRGVGQRLGE
GESQRYMYRMPDPVMGVCEMVYRRLFGESSNMAPYRDMNVRSSRRRAMQAEESLSRVYL
FLLCFMFMLFLF*

>G597 (255..1310)

AAAATTCTCTGTAAATTTAATATTATAAAAGTGGTTTCTTTTTCAATTTATGTTTATAT
AATTTTCATCTTTAATCTTAAATTCTGGTAACCTTAATGCGCGATCCGCTTTTCTAAAGT
TTTGTGAGAGAGAAGAGATCTAAAAAATCCACAATTTTGTTCAAATCTTGAGGTTAAAT
GCTGAATTTTAGGCCTTGTGCTTAGATTTATGGCTTAAAGTTTCAAACTTTTCATTGGA
TATGTGAGAAGAAATGTCAGGATCTGAGACGGGTTTAAATGGCGCGACCAAGAGAATCAA

TGCAATTTACAATGGCTCTCCACCAGCAGCAGCAACACAGTCAAGCTCAACCTCAGCAGT
CTCAGAACAGGCCATTGTCTATTCGGTGGAGACGACGGAAGTCTCTTTACAAGCAGCCGA
TGAGATCAGTATCACCACCGCAGCAGTACCAACCCAAGTCAAGCTGGTGAGAATTCTGTCT
TGAACATGAAGTTGCCCGGAGGTGAGTCTGGAGGCATGACTGGAAGTGAAGTGAGCCAG
TGAAAAGAGGAGAGGTAGACCGAGGAAATATGGGCCTGATAGTGGTGAAATGTCACTTG
GTTTGAATCCTGGAGCTCCTTCTTTCACTGTCTAGCCAACTAGTAGCGGCGGCATGGAG
GAGAGAAGAAGAGAGGAAGACCTCCTGGTTCTTCTAGCAAAAGGCTCAAGCTTCAAGCTT
TAGGCTCGACTGGAATCGGATTTACGCCTCATGTACTTACCGTGTGGCTGGAGAGGATG
TATCATCCAAGATAATGGCGTTAACTCATAATGGACCCCGTGTGTGTGTCTTGTCTG
CAAATGGAGCCATCTCCAATGTGACTCTCCGCCAGTCTGCCACATCCGGTGAAGTGTTA
CATATGAGGGGAGATTGAGATTCTGTCTTTATCGGGATCTTTCCATTTGCTGGAGAACA
ATGGTCAAAGAAGCAGGACGGGAGGTCTAAGCGTGTCAATTATCAAGTCCGGATGGTAATG
TCCTCGGTGGCAGTGTAGCTGGTCTTCTTATAGCAGCATCACCTGTTTCAAGTGTGTG
GGAGTTTCTTACCAGACGGAGAAAAGAACCACCAACAGCATGTGGGACAAATGGGACTGT
CGTCACCCGTATTACCGCGTGTGGCCCCAACGCAGGTGCTGATGACTCCAAGTAGCCAC
AATCTCGAGGCACAATGAGTGAGTCATCTTGTGGAGGAGACATGGAAGCCCTATTTCATC
AGAGCACTGGAGGACCTTACAATAACACCATTAACTATGCCCTGGAAGTAGCCAAGTGATC
TGTGTGCGCTTAAACCAACAACCTTCCCGTTATTAGAGTGATTTATTCTACATTTGGTT
TAGACTTTCTAGTTCTGATGGTTATTTCTACAGTTGGTTTAGACTTTCTAGTTCTGTTCA
GACAAAAGGAGTTTGATAAATTGACCGACCTATTTTGTGTGTTTGAGGTACTTTTCAGAAC
CATAGGTGTTTCAGAAATTAGAATGTTCTGTTTAAAAAA
>G597 Amino Acid Sequence (domain in AA coordinates: 97-104,137-144)
MSGSETGLMAATRESMQFTMALHQQQHSQAQPQSQNRPLSFGDDGTALYKQPMRSVS
PPQQYQPNASAGENSVLNMLPGGESGGMTGTGSEPVKKRRGRPRKYGPSGEMSLGLNPG
APSFTVQPSGGDGGGKKRRGRPPGSSSKRLKLQALGSTGIGFTPHVLTVLAGEDVSSKI
MALTHNGPRAVCVLSANGAISNVTLRQSATSGGTVTYEGRFEILSLSGSFHLLNNGQRS
RTGGLSVSLSSPDGNVLGGSVAGLLIAASPVQIVVGSFLPDGEKEPKQHVGMGLSSPVL
PRVAPTQVLMTPSSPQSRGTMSSESSCGGHHGSPIHQSTGGPYNNTINMPWK*
>G1009 (28..1704)
AAAAAAAAAAAAACCTATTCCCAAAGATGAAGAACAATAACAACAAATCTTCTTCTTCT
TCTAGCTATGATTCTTCTTTGTCTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
TCTTTCTTCTTCTTCCAAACAATAACAACAACTTCAATTCTTCTTCTTCTTCTTCTTCTTCT
TCCTCCACATCAGATCATCATCATCTTCCACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
TCCACTTCTCCAGTCGAACGGCAAGATGGGTACCGGGAGTTTACCCAGCGATGCCACG
GCGGTTCTTCCGTATACCCCGCGGTCTTAACTTGAAGCTTCTTCTTCTTCTTCTTCTTCTTCT
TCAACGACGACAACAAGCAATGCAACAAGTGCAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
TCCGACCTACAGCCACCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
AAGTCAATAGCCGCTAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
AGCGTACATAAACAACACCGAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
AAGAACGTAGAGAGTTTGGACAACGTACCTCGATTATATAGAGGAGTCAACAAGACATAGA
TGGACTGGAAGATACGAAGCTCATCTATGGGATAATAGTTGCCGAAGAGAAGGCCAAAGC
AGAAAAGGAAGACAAGTTATTTAGGTGGTTATGATAAGGAAGATAAAGCACTAGAGCT
TACGACCTTGCAGCTCTTAAAGTATTGGGGTCTTCAACTACGACTAATTTCCCGATATCA
AATTACGAATCTGAAGTTGAAGAAATGAAACACATGACTCGACAAGAGTTTCGTTGCTTCT
TTAAGACGGAAGAGCAGTGGATTCTTCTAGGGGTGCCTCCATGTACAGAGGCGTCACTAGA
CATCATCAGCATGGTTCGATGGCAGGCACGAATTGGAAGAGTTGCAGGCAACAAAGACCTT
TATCTTGGCACATTTAGCACTCAAGAGGAAGCTGCAGAAGCTTATGATATAGCAGCGATC
AAATTCGCGGTCTAATGCAGTCACCAATTTTCGACATCAGTCGATATGATGTCAAATCA
ATTGCTAGCTGTAATCTCCTGTGGGTGGACTAATGCCCTAAACCTTCTTCCAGCAACCGCA
GCGGCTGACAAAACCGTTGATCTTTCTCCATCCGACTCTCCATCTCTAACCACACCGTCC
CTCACGTTCAATGTGGCAACACCGGTCAATGACCATGGAGGAACTTTTTACCACACTGGT
ATACCAATCAAACAGACCCGGCTGATCATTATTGGTCCAACATCTTTGGATTCCAGGCA
AACCCGAAAGCAGAAATGCGACCATTAGCAAACTTTGGGTTCGGATCTTCTAACCCTTCT
CCTGGTTATGCTATAATGCCGGTAATGCAGGAAGGTGAAAACAACCTTTGGTGGTAGTTTT
GTTGGGTCTGATGGGTATAACAATCATTCGCTGCATCGAACCCGGTCTCAGCAATCCG
CTGTCTCGACAACCTACATGAGTAACGGTAACGAAGGGTATGGTGGAAACATAAACTGG
ATTAATAACAACATTTCAAGTTCTTACCAAAAGTCAAAATCAATCTCTGTTTTGCAC

ACACCGGTTTTTGGGTTGGAATGAGTATTCACATCTTAGTGAGAACTAAAATAAATATGT
AGGAAAAAATAAGGCTCTGTTTGAAGAAATCAGATATTTCTTCTTAGATTATTTAAGT
AGTTTAAAAAATATTTTTTAAAGTGTTTCACCTTTACGTTTGTCTGCTGACCACGAATT
TTGCTGGATCTGACAGTACTAAGTCTTTGTTTAAATGACCTTATGGGTTCCCTTTTTTACTT
TCCAGAACTTTTATTTACTTTTTTCTTCATTTTTCTTCATTTTTTTTGTGTGGGACAAT
ATGAATGATTGAAGATGGAACGCTTGCATGTGAATAAACGAAAATCAAACNATCTTCG
GTAACTTAAAAA

>G1009 Amino Acid Sequence (domain in aa coordinates: 201-277, 303-371)

MKNNNNKSSSSSYDSSLSPSSSSSHQNWLSFSLNNNNNFNSSSNPNLTSSTSDHHHP
HPSHLSLFQAFSTSPVERQDGS PGVSPSDATAVLSVYPGGPKLENFLGGGASTTTTRPMQ
QVQSLGGVVFSSDLQPPPLHPPSAAEIYDSELKSI AASFLGNYSGGHSSEVSVHKQQPNP
LAVSEASPTPKKNVESFGQRTSIYRGVTRHRWTGRYEHLWDNSCRREGQSRKGRQVYLG
GYDKEDKAARAYDLAALKYWGPTTTTNFPISNYESELEEMKHMTRQEFVASLRRKSSGFS
RGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFTQEEAAEAYDIAAIKFRGLNAV
NFDISRYDVKSIAASCNLPVGGMLPKPSPATAAADKTVDLSPSDSPSLTTPSLTFNVATPV
NDHGGTFYHTGIPIKPD PADHYWSNIFGFQANPKAEMRPLANFGSDLHNPSPGYAIMPVM
QEGENNFGGSFVGS DGYNNHSAASNVPVSAIPLSSTTTMSGNGEGYGGNINWINNNISSY
QTAKSNLSVLHTPVFGL*

>G170 (1..1107)

ATGGGGATGAAGAAGGTGAAGCTATCTTTGATAGCTAATGAAAGATCAAGGAAAACATCC
TTCATAAAGAGGAAAGACGGGATTTTAAAGAACTCCACGAGTTGTCAACTCTGTGTGGT
GTCCAAGCTTGTGCTCTCATCTACAGTCCATTCATAACGGTTCAGAGTCATGGCCGTCA
AGGGAAGGTGCTAAAAAGGTGGCTTCAAGGTTTCTGGAGATGCCGCCGACAGCCCGAACC
AAGAAGATGATGCTCAAGAGACTTACCTTATGGAGAGGATTACCAAAGCAAAGAGCAA
CTAAAGAACCCTGGCTGCTGAGAACCGAGAGTTACAGGTTAGACGATTATGTTTGTATTGT
GTTGAAGGC AAAATGTCCCAGTATCATTATGATGCAAAAGACCTTCAAGATTTGCAATCT
TGTATAAATCTATATCTCGATCAGCTTAACGGAAGGATCGAGTCCATTAAAGAAAATGGT
GAGTCGTTGTTGTCTTCCGTCTCTCCTTTTCTTACTAGAAATTGGTGTGACGAAATTGGT
GATGAGTCATTTTCCGACTCTCCTATTATGCTACAACCTGGGTTGTAGATACTCTTAAT
GCTACCAATCCTCATGTTCTTACGGGCGATATGACTCCTTTTCTTGATGCGGACGCAACT
GCGGTAAGTGTCTCCAGTAGATTTTGTGATCATATCCATATGAAAATATGAATATGAGT
CAAAATCTGCATGAACCGTTTCAACACCTTGTCTTACTAACGTTTGTGATTTTTTTCAA
AATCAGAATATGAATCAGGTTCAATACCAGGCTCCTAATAATCTGTTTAAATCAGATTCAA
CGAGAATTCTACAACATAAATTGAATCTGAATTTGAATCTGAATTCGAATCAGTATCTG
AATCAACAACATCATTATCATGAATCCGATGTTGGAACAACATATGAATCATGTTGGAGGG
CGTGAAAGCATTCTTTCTGTTGACGGAACTGCTACAACCTACCATCAATACCATCCAAT
CAACTACGAGCGTTGATCATGCTTCCACCAGTTACATGCCTTCCACCACCGGTGTCTAT
GATCCTTACATCAACAATAATCTCTAA

>G170 Amino Acid Sequence (domain in aa coordinates: 2-57)

MGMKKVKLSLIANERSRKTSPFKRKDGIFKKLHELSTLCGVQACALIYSPFIPVPESWPS
REGAKKVASRFLEMPPTARTKKMMDQETYLMEITKAKEQLKNLAAENRELQVRRFMFDC
VEGKMSQYHYDAKDLQDLQSCINLYLDQLNGRIE SIKENGESLLSSVSPFPTRIGVDEIG
DESFSDSPIHATTGVVDTLNATNPHVLTGDMTPFLDADATAVTASSRFPDHIPIYENMMS
QNLHEFPQHLVPTNVCDFFQONQNMNQVQYQAPNNLFNQIQREFYINLNLNLNSNQYL
NQQSFNMNPMVBEQHMNHVGGRESIPFVDGNCYNYHQLPSNQLPAVDHASTSYMPSTTG
VYDPYINNLL*

>G1768 (185..1426)

CTTCCTTTTGCTTCAGCTGCGAGCTTTGGTTGGATCTCTCACTTGCAAAACCAAATCCCT
TATCGACTTCCACCGAAAGATCACTTCTTAACCTACACAAGGTGTTTGTATGAAGATCA
GATAAATAAAGGTCAATTGAGGATAATGGTTGATGTTCAAAGATTCTTACTTGCTTATT
TGTGATGGCAATGTAAGAGGTTCAATAATGTTGCAGCCACTGCCAGAGATAGCTGAGAG
TATCGATGATGCTATCTGCCATGAACCTCTCCATGTGGCCTGATGATGCTAAAGATTGTT
ATTGATAGTGGAGGCAATATCAAGGGGAGACTTGAAGTTGGTACTTGTGCTTGTGCAAA
AGCTGTTTCTGAGAATAATCTTCTAATGGCACGATGGTGTATGGGTGAGTTGCGCGGTAT
GGTTTTCGATTTCTGGTGAGCCAATCCAGAGATTGGGAGCTTATATGTTAGAAGGGCTGT
TGCTAGGCTTGCTGCTTCTGGTAGTTCGATATATAAGTCTCTCCAGTCCAGAGAACCAGA
GAGTTATGAATTTTATCTTATGTGTATGTTCTGCATGAGGTTTGTCCATATTTCAAGTT

TGGATACATGTCAGCGAATGGTGGCATTGCAGAAGCAATGAAGGATGAAGAGAGGATTCA
CATTATTGACTTCCAAATTTGGACAAGGGAGCCAGTGGATAGCATTATCCAGGCTTTTGC
AGCTAGGCCTGGTGGGGCTCCAAATATTGCAATTACCGGAGTTGGTGATGGATCTGTCTT
GGTTACAGTCAAGAAGAGACTAGAGAACTTGCAAAGAAGTTTGATGTTCCATTACAGGT
CAATGCGGTTTCAAGGCCAAGTTGTGAAGTTGAAGTGGAAATCTTGATGTCCGAGATGG
CGAAGCCCTTGGAGTGAACCTTTGCTTACATGCTGCATCATTGTCAGATGAGAGTGAAG
CATGGAAAACACAGGGACCGGTTGCTGAGGATGGTGAAGAGTCTATCACCTAAAGTAGT
CACTCTTGTTGGAACAAGAATGCAACACGAACACTTCCCTTTCTCTTAGGTTCTTGA
GACATTAAGTTATTACACGGCAATGTTGCAATCTATCGATGTTATGCTTCCGAGAAATCA
CAAGGAAAGGATCAATATCGAGCAGCACTGCATGGCAAGGGATGTCGTCAACATCATAGC
TTGTGAAGGAGCCGAGAGGATCGAAAGACACGAGCTTCTCGGGAATGGAAGTCAAGGTT
TTCCATGGCGGGTTTGGAGCCATACCCCTTGAGCTCAATCATTTCAGCCACCATTAGAGC
CCTCTTGAGAGATTACAGCAACGGGTATGCGATTGAAGAAAGAGATGGTGCTCTGTACCT
TGGTTGGATGGACCGAATCTTGGTCTCATCTTGTGTCATGGAAGTGAAGAATAAACGCTCT
CCAAGAATGTAATGCAAAAGACAGAACTGGAAGTAATAGATAGTTTGTCTCATAACCAT
TAATAAGGTTGAATCAAATCATATACATCCCATGCTACAACCTATTACACAGGCTCCATC
AACAAAGAGGGCTCTTGTGTGTACCTTCTCTCTGTAACCTTATTTGAACCAAAT
GGAAGTGGTTACAT

>G1768 Amino Acid Sequence (domain in AA coordinates: 54-413)
MDNVRGSI MLQPLPEIAESIDDAICHEL SMWPDDAKDLLLIVEAISRGDLKLVLVACAKA
VSENNLLMARWCMGELRGMVSI SGEPIQRLGAYMLEGLVARLAASGSSIIYKSLQSRPES
YEFLSYVYVLHEVCYPYFKFGYMSANGAIAEAMKDEERIHIIDFQIGQGSQWIALIQAF
RPGGAPNIRITGVGDGSLVLT VTKRLEKLAKKFDV PFRFNAVSRPSC EVEVENLDVRDGE
ALGVNFAYMLHLLPDESVMENHRDRLRMVKSLSPKVVT LVEQECNNTS PFLPRFLET
LSYYTAMFESIDVMLPRNHKERINIEQHCMARDVVNIIACEGAERIERHEL LGKWSRFS
MAGFEPYPLSSII SATIRALLRDYSNGYAI EERD GALYLGMMDRILVSSCAWK*

>G185 (77..988)

ATGCAAAAATAAACATAGTAACAATACTTTAAACTATTTACACCACCTTTAATCTTATTCT
CCACTCTTTGAACGTAATGGAGAAGAACCATAGTAGTGGAGAGTGGGAGAAGATGAAGAA
CGAGATCAACGAGCTAATGATAGAAGGAAGAGACTATGCACACCAGTTTGGATCAGCTTC
ATCTCAAGAAACACGTGAACATTTAGCCAAAAAGATTCTTCAATCTTACCACAAGTCTCT
CACCATCATGAACACTCTCCGGCGAAGTTGACCAAGTTTCTCAGGGTGGAGGAAGCCCCAA
GAGCGATGATTCCGATCAAGAACCACTTGTCATCAAGAGTTCGAAGAAGTCAATGCCAAG
GTGGAGTTCAAAAGTCAGAAATGCCCCTGGAGCTGGTGTGATAGAACGCTGGACGATGG
ATTGAGTTGGAGAAAGTACGGCCAGAGGATATTCTCGGAGCCAAATTTCCAAGAGGATA
CTATAGATGCACGTATAGAAAGTCTCAAGGATGTGAAGCCACTAAACAAGTCCAAAGATC
TGATGAAAATCAGATGCTCCTTGAGATCAGTTACCGAGGAATACATTCTTGCTCTCAAGC
TGCAAATGTCGGTACAACAATGCCGATACAAAACCTCGAACCAGAACCCAGACCCAAAGACA
CGGAAATCTTGACATGGTAAAGGAAAGTGTAGACAACCTACAATCACCAGCACATTTGCA
TCACAACCTTCACTATCCATTGTCTATCTACCCCAATCTAGAGAATAACAATGCCTATAT
GCTTCAAAATGCGAGATCAAAACATCGAATATTTTGGATCTACGAGCTTCTCTAGTGATCT
AGGAACTAGTATCAACTACAATTTTCCAGCATCTGGCTCGGCTTCTCACTCAGCATCAAA
CTCTCCGTCCACCGTCCCTTTGGAATCCCCGTTTGAAGCTATGATCCAAATCATCCATA
TGGAGGATTTGGTGGGTTCTATTCTTAGTTATCTACTTAAGGGAGGGACGGAACTTTTTA
CATGACCTCTTGATTAAAGAGAGAGTTTTCATAATAGCTAATCAATTTCTTATTCAAATA
TCCGAGTTTTTTTTCTAATCATGTTTATCAATTGTCTTATTACAGAAGGCTTATTTTCAG
GTCTATGTTGAAATAAATGGATTTGTACTCGTAGGTATGATCCTTGTATCTAAAAA
AAAAA

>G185 Amino Acid Sequence (domain in AA coordinates: 113-172)
MEKNHSSGEWEKMKNEINELMIEGRDYAHQFGSASSQETREHLAKKILQSYHKS LTIMNY
SGELDQVSQGGGSPKSDSDQEPLVIKSSKKSMPRWSSKVRIAPGAGVDR TLDDGF SWRK
YGQKDILGAKFPRGYRCTYRK SQGCEATKQVQRSDENQMLLEISYRG IHS CSQAANVGT
TMPIONLEPNQTQEHGNLDMVKESVDNYNHQAHLHNLHYPLSSTPNLENNNAYMLQMRD
QNIEYFGSTSFSSDLGTSINYNFPASGSASHSASNSPSTVPLES PFESYDPNHPYGGFGG
FYS*

>G1931 (5..592)

ATCAATGGAAGGGGTTGACAACACAAATCCTATGTTAACCCTAGAAGAAGGCGAAAAACAA

CAATCCTTTTCTTCTTAGATGACAAAACATTAATGATGATGGCTCCTTCGTTAATCTT
TTCGGGCGATGTAGGTCCATCTTCTTCTTCTGTACTCCAGCAGGTTATCATCTATCTGC
TCAGCTGGAGAACTTTCGAGGAGGTGGAGGAGAGATGGGAGGATTAGTGAGTAATAATAG
CAATAATAGTGATCATAATAAGAATTGCAACAAAGGAAAAGGGAAGAGAACTTTGGCAAT
GCAGAGGATAGCTTTTCATACAAGGAGTGATGATGATGTTCTTGATGATGGTTATCGTTG
GCGAAAGTACGGTCAGAAATCTGTCAAGAACAATGCTCATCCCAGGAGCTATTATAGATG
TACATACCACATGCAACGTGAAGAAACAAGTGCAAGACTGGCAAAAGATCCAAACGT
TGTCGTAACAACCTACGAAGGTGTTTCATAATCATCCTTGTGAGAAGCTCATGGAGACTCT
TAGCCCTCTCCTTAGGCAACTTCAGTTCCTCTCAAGAGTTTCTGATCTGTAATTATTGAA
TGTTAATTAGTGGTGTAATACATTAATTATGCTTTAATCTCTCCATTGACCCCTCAATC
>G1931 Amino Acid Sequence (domain in AA coordinates: 114-170)
MEGVDNTNPMLEEGENNPFSSLDKTLMMAPSLIFSGDVGPSSSSCTPAGYHLSAQ
LENFRGGGEMGGLVSNNSNNDHNKNCNKGKGRITLAMQRIAFHTRSDDDVLDGYSRWR
KYGQKSVKNNAHPRSYRCTYHTCNVKKQVORLAKDPNVVVTYEGVHNHPCEKLMETLS
PLLRLQLFLSRVSDL*

>G2543 (1..2169)
ATGAGTTTCGTCGTCGGCGGAGGTGGTAGTGGAAGCGGCGGAGACGGTGGTGGT
AGTCATCATCAGCAGCGCTCTGAACTGATAGGAAGAAGAAACGTTACCATCGTCACACC
GCTCAACAGATTCAACGCTTGAATCGAGTTTCAAGGAGTGCTCTCATCCAGATGAGAAA
CAGAGGAACAGCTTAGCAGAGAATTGGGTTTGGCTCCAAGACAAATCAAGTTCTGGTTT
CAGAACAGAAAGAACTCAGCTTAAGGCTCAACATGAGAGAGCAGATAATAGTGCATAAAG
GCAGAGAATGATAAAATTCGTTGCGAAAACATTGCTATTAGAGAAGCTCTCAAGCATGCT
ATATGTCCTAACTGTGGAGGTCTCTCTGTTAGTGAAGATCCTTACTTTGATGAACAAAAG
CTTCGGATTGAAAATGCACACCTTAGAGAAGAGCTTGAAAGAATGTCTACCATTGCATCA
AAGTACATGGGAAGACCGATATCGCAACTCTCTACGCTACATCCAATGCACATCTCACCG
TTGGATTGTCAATGACTAGTTTAACTGGTTGTGGACCTTTTGGTCATGGTCTCTCACTC
GATTTTGTATCTTCTCCAGGAAGTTCTATGGCTGTTGGTCTAATAATAATCTGCAATCT
CAGCCTAACTTGGCTATATCAGACATGGATAAGCCTATTATGACCGGCATTGCTTTGACT
GCAATGGAAGAATTGCTCAGGCTTCTTCAGACAAATGAACCTCTATGGACAAGAACAGAT
GGCTGCAGAGACATTCTCAATCTTGGTAGCTATGAGAATGTTTTCCCAAGATCAAGTAAC
CGAGGGAAGAACCAGAACTTCGAGTCGAAGCATCAAGGTCTTCTGGTATTGCTTTCATG
AATGCTATGGCACTTGTGACATGTTTCATGGATTGTGTCAAGTGGACAGAACTCTTCCC
TCTATCATTGCAGCTTCTAAAACACTTGCAGTGATTCTTTCAGGAATGGGAGGTACCCAT
GAGGGTGCATTGCATTTGTTGTATGAAGAAATGGAAGTGCTTTCGCCTTTAGTAGCAACA
CGCGAATTCTGCGAGCTACGCTATTGTCAACAGACTGAACAAGGAAGCTGGATAGTTGTA
AACGCTCTCATATGATCTTCTCAGTTTGTCTTCTCACTCTCAGTCTTATAGATTTCCATCT
GGATGCTTGATTAGGATATGCCCCAATGGATATTTCAAGGTTACTTGGGTTGAACATATT
GAAACTGAAGAAAAGAACTGTTTCATGAGCTATACAGAGAGATTATTCACAGAGGGATT
GCTTTTGGGGCTGATCGTTGGGTTTACCACTCTCCAGAGAATGTGTGAAAGATTGCTTCT
CTATCGGTACCAGCGTCTTCATCTCGTGATCTCGGTGGAGTGATTCTATCACCAGGAGGG
AAGAGAAGCATGATGAGACTTGCTCAGAGGATGATCAGCAACTACTGTTTAAAGTGCAGC
AGATCCACAACACACGCTCAACCGTTGTTTCGGAAGTGAACGAAGTTGGAATCCGTGTG
ACTGCACATAAGAGCCCTGAACCAACCGGCACAGTCTATGTGCAGCCACCACCTTCTGG
CTTCCCAATTCTCTCAAAATGTCTTCAATTTCTCAAGACGAAAGAACCCGTCCTCAG
TGGGATGTTCTTTTCAAACGGAACGAGTGCAAGAAGTTGCTCACATCTCAAACGGATCA
CATCCTGGAACTGCATATCGGTTCTACGTGGATCCAATGCAACACATAGCAACAACATG
CTTATTCTGCAAGAAAGCTCAACAGACTCATCAGGAGCATTTGTGGTCTACAGTCCAGTG
GATTTAGCAGATTGAACATCGCAATGAGCGGTGAAGATCCTTCTTATATTCTCTCTTG
TCCTCAGGTTTCAATCTCACCAGATGGAAATGGCTCAAACCTGAAACAAGGAGGAGCC
TCGACGAGCTCAGGACGGGCATCAGCTAGCGGTTGCTTGATAACGGTTGGGTTTCAGATA
ATGGTAAGCAATTTACCGACGGCAAACTGAATATGGAGTGGTGGAAACGGTTAATAAC
CTGATAGGAACAACCTGTACATCAAATTAACCCGCTTGAGCGGTCTACAGCTTCAACT
ACAGCTTGA

>G2543 Amino Acid Sequence (domain in AA coordinates: 31-91)
MSFVVGSGSGSGSGDGGGSHHHDGSETDRKKRYHRHTAQQIQRLESSFKECPHPDEK
QRNQLSRELGLAPRIKFWFQNRRTQLKAQHERADNSALKAEENDKIRCENIAIREALKHA
ICPNCGGPPVSEDYPFDEQKLRIENAHLEBLERMSTIASKYMGRPISQLSTLHPMHISP

LDLSMTSLTGCGPFGHGPSLDFDLLPGSSMAVGPNNNLQSQPNLAISDMDKPIMTGIALT
AMEELLRLQLTNEPLWTRTDGCRDILNLGSYENVFPRSSNRGKNQNFVEASRSSGIVFM
NAMALVDMFMDCKVWTELPFSIIAASKTLAVISSGMGGTHEGALHLLYEEMEVLSPLVAT
REFCELRYCQQTEQGSWIVNVSYDLPQFVSHSQSYRFPSPGCLIQDMPNGYSKVTWVEHI
ETEEKELVHELRYEIIHRGIAFGADRWVTTLQRMCFERFASLSVPASSSRDLGGVILSPEG
KRSMMLRAQRMISNYCLSVSRNNRSTVSELNEVGIRVTAHKSPEPNGTVLCAATTFW
LPNSPQNVFNFLKDERTRPQWDVLSNGNAVQEVAHISNGSHPGNCISVLRGSNATHSNNM
LILQESSTDSSGAFVVPDLAALNIAMSGEDPSYIPLLSSGFTISP DGNGSNSEQGGA
STSSGRASASGLITVGFIQIMVSNLPTAKLNMESVETVNNLIGTTVHQIKTALSGPTAST
TA*

>G264 (30..1430)

CTTGTACAGTTTCTGATTAGATTCAACAATGAACGGCGCATTAGGTAACTCCTCCGCCT
CCGTTAGCGGCGGAGAAGGAGCCGAGGACCAGCGCCTTTCTTGGTGAAAACCTACGAGA
TGGTCGACGATTCATCAACGGACAGATCGTATCGTGGAGCGCTAACAAACAGCTTCA
TCGTTTGGAAATCATGCCGAATTTTCACGCCTCCTTCTTCCAACCTACTTCAAACACAATA
ACTTCTCTTCCTTCATTCTCGTCAGCTCAATACCTATGGGTTTAGGAAGATTGATCCAGAGA
GGTGGGAGTTTTTGAATGATGATTTTATTAAGGATCAGAAGCATCTTCTCAAGAATATAC
ATAGAAGGAAACCTATACACAGCCACAGTCATCCACCTGCTTCGTGCGACTGATCAAGAAA
GAGCAGTGTGCAAGAGCAAATGGACAAGCTTTCACGTGAGAAAGCTGCAATTGAAGCTA
AGCTTTTAAAGTTCAAACAACAGAAGGTTGTAGCAAAGCATCAGTTTGAAGAAATGACTG
AGCATGTTGATGATATGGGAATAGGCGAGAAGAAGCTGCTGAATTTTTTGGAAACTGCGA
TTCGGAATCCTACTTTTGTAAAGAAATTTTGGTAAGAAAGTCGAGCAGTTGGATATTTTCA
CTTACAACAAAAAGCGAAGGCTCCCTGAAGTTGAGCAATCAAAGCCACCTTCAGAAGATT
CTCATCTGGATAATAGTAGTGGTAGCTCGAGACGCGAGTCTGGAAACATTTTTTCATCAA
ATTTCTCTAATAAATTTGCGACTAGAGCTTTCTCCAGCTGATTGAGATATGAACATGGTTT
CACACAGTATACAAAGTTCCAATGAAGAAGGTGCGAGTCCCAAAGGGATACTGTCAAGGAG
GTGATCCAAATACTACACTAACAAAAAGAGAAGGCTTACCATTGTCACCTGAAGCTCTAG
AGCTTGCGGATACCGGGACATGCCCGAGGAGATTACTGTAAATGATAATAACAAGGGTGG
AGACCTTGCGAGCAGAGGCTAACTTCTTCAGAGGAGACTGATGGTAGCTTTTCATGTCTATT
TAAATCTAACCTTGGCTTCTGTCTCCGTTACCGGACAAAACAGCTTCACAGATAGCTAAGA
CGACTCTTAAAGTCAAGGATTAACCTTTAACTCAATAGAAACAAGTGAAGTGAGAAAA
ATCGGGGTAGACAAGAGATTGCAGTTGGAGGTAGCCAAAGCAAATGCAGCTCCTCCAGCAA
GAGTGAATGATGATTCTGGGAACAGTTCTTAACAGAAAGGCCAGGGTCTTCAGATAATG
AGGAGGCAAGTTCCGACTATAGAGGTAACCCATACGAAGAGCAAGAGGAGAAAAAGAAACG
GGAGTATGATGTTTACGTAATACAAAGAATATCGAGCAGCTGACCTTATAAACTATTTTGGA
CGGTTACATCAACGAGAGTACGAAGTGGGTTTTGGTAAGAAAGTATGGGTGAGTAAGTAA
TGAAACATTGGACTGAAAAAGCGTAAGTAGCTTTGTTGTAAACACTTGCGTCTCTGTCTA
CACAAGTAATTGACTGTAAATGTAAGTGTACAGGATTAAATTGAATAAGCA

>G264 Amino Acid Sequence (domain in AA coordinates: 24-114)

MNGALGNSSASVSGGEGAGGPAPFLVKTYEMVDDSSDQIVSWSANNNFSFIVWNHAEFSR
LLLPTYFKHNNFSSFIQNLNTYGFIRKIDPERWEFLNDDFIKDQKHLKNIHRRKPIHSHS
HPPASSTDQERAVLQEQMDKLSREKAAIEAKLLKFKQKQKVVAHQFEEMTEHVDDMENRQ
KKLLNFLETAIRNPTFVKNFGKKVEQLDISAYNKKRRLPEVEQSKPPSEDSHLDNSSGSS
RRESGNIFHQNFNKLRLLELSPADSDMNMVSHSIQSSNEEGASPKGILSGGDPNTTLTKR
EGLPFAPEALELADTGTCPRRLLLNNDNTRVETLQQLTSSSEETDGSFCHLNLTLASAPL
PDKTASQIAKTTLSQELNFSIETSASEKNRGRQEIIVGGSQANAAPPARVNDVFWEQF
LTERPGSSDNEEASSTYRGNPYEEQEERNGSMMLRNNTKNIEQLTL*

>G32 (101..736)

AACACACATTCCTCTCTCTTCTCAACTAGAAAAAGATAGATATATCGGACATTTATTG
ATCTGTGTATGCATAAAGGTATAGTATCATTATTAGAAAGATGAACACAACATCATCAA
GAGCAAGAAGAAGCAAGACGATCAGGTTGGTACAAGGTTTCTTGGGGTGAGAAGAAGGCC
TTGGGGAAGATACGCAGCTGAGATTAGAGACCCAACTACGAAGGAGCGTCACTGGCTTGG
CACTTTTCGATACGGCGGAAGAAGCTGCCTTGGCCTACGATAGAGCTGCTCGGTCCATGCG
TGGCACACGTGCCAGAACCAACTTTGTTTACTCAGACATGCCTCCTTCTCATCCGTCAC
CTCCATTGTTTCTCTCAGCATCCTCCTCCTCCACCTCCTCTCTCTCTCTCTAGCAA
TGATCCTGTCTACATGATGATGTTTAAACAACTACTCATCCACTGACTCGCCAATGCT
TCAGCCTCATTGTGATCAAGTGGACAGTTACATGTTTGGTGGCTCTCAATCTTCGAATTC

TTATTGCTATTCTAATGACAGTAGTAATGAGCTGCCTCCTCTCCCGAGCGACTTGTGCGAA
TTCGTGTTATAGCCAACCACAGTGGACCTGGACCGGTGACGACTACTCGTCTGAGTACGT
ACATAGTCCAAATGTTTCAGCAGAATGCCTCCGTTTCTGACTCTTCCCTCAAGGTTTCAA
CTACTTTGGCTCCTAATCTTTCTCATCGTCCATATTTAATACCTTCTCATTTGTACCT
TTTCCTTCTTCTTCTTTTTTGGGTTTATCTATGTTTCGCCGTCTTGATCTCTGCCTATG
TGATCAAAGTGACTGTTTGTCAATAGTTTTTCAATAACAAGTTATCATTTGTATCTTGAA
AAAAAAAAAAAA

>G32 Amino Acid Sequence (domain in aa coordinates: 17-84)
MNTTSSKSKKKQDDQVGTRFLGVRRRPWGRYAAEIRDPTTKERHWLGTFDTAEAAALAYD
RAARSMRGTRARTNFVYSMDMPSSSVTSIVSPDDPPPPPPPPAPPSPNDPVDYMMMFNQYS
STDSPMLQPHCDQVDSYMFSGSQSSNSYCYSDSSNELPPLPSDLSNSCYSPQOWTWTGD
DYSSEYVHSPMFSRMPVSDSFPQGFNYFGS*

>G436 (1..2157)

ATGGATTTTACTCGCGATGACAACTCAAGTGATGAACGGGAAAATGATGTAGACGCCAAC
ACCAACAACCGTCACGAGAAGAAGGGTTACCATCGCCACACTAATGAACAAATTCATAGG
CTTGAAACGTATTTCAAGGAATGTCTCATCCAGACGAATTCAGCGACGTCTGTTGGGT
GAAGAAGCTGAATCTGAAACCAAAAACAAATCAAATTTTGGTTTCAAACAAAAGAAGTCAA
GCTAAGAGTCACAATGAAAAAGCAGACAATGCAGCGCTTAGGGCAGAAAATATTAAAGATT
AGACGTGAGAACGAATCAATGGAAGATGCACTGAATAATGTGGTTTGCCCTCCATGTGGT
GGTTCGTGGTCTCGGAGAGAGACCAACTTCGACATCTCCAAAACCTCCGTGCACAAAAC
GCTTATCTCAAAGATGAGTATGAAAGAGTCTCAAACACTCTAAAACAGTACGGAGGTCAC
TCAATGCATAACGTCGAGGCCACACCTATCTCCATGGTCCATCAAACCATGCATCAACG
TCCAAGAACCGTCAGCATTGTACGGAACCTCTTCTAACCGTCTCCCGAGCCTTCAAGC
ATATTTAGAGGACCATACACTCGTGGAACATGAACACCACCGCACCGCTCAGCCGCGA
AAGCCGCTGGAATGCAGAAATTTCCAACCACTATCTCAACTGGAGAAAATTGCAATGTTG
GAAGCAGCGGAAAAAGCGGTGTGAGAGGTTTGGAGCCTCATTCAAATGGATGATACAATG
TGGAAAAAGTCGTCTATTGATGATAGGCTCGTCATTGATCCAGGGCTCTATGAGAAATAT
TTTACTAAGACTAACACAAATGGTCGTCTGAGTCTTCTAAAGATGTCTGGTGGTTCAA
ATGGATGCTGGAACTTGATCGACATCTTCTTAACTGCGGAGAAATGGGCGAGGCTTTTT
CCAACAATTGTGAACGAAGCTAAAACGATTACGCTCTTGGATTCCGTTGACCATCGAGGA
AAAACCTTTCTCAAGAGTGATTTATGAGCAACTGCACATACTGTACCATTTGGTGCCACCG
AGGGAATTTATGATCCTAAGGACTTGCCAACAAATGAAGACAATGTCTGGATGATTGCT
GATGTGTGCTGTCTATCTCCCAAACATTGAGTTTGATCTTTCGTTTCCCATTTGCACCAAA
CGTCCCTCAGGTGTGCTCATTCAGCCTTGCCCCACGGCTTCTCTAAGGTGACGTGGATA
GAGCATGTGGTAGTGAATGATAATAGAGTGCGGCCACATAAGCTTTACAGAGACCTCTTA
TACGGCGGCTTTGGCTACGGAGCTCGACGTTGGACCGTTACTCTTGAGAGGACGTGTGAG
AGGCTGATTTTCTCCACCTCCGTCCCTTGCCCAACAATGACAATCCCGGAGTTGTG
CAAACAATACGAGGCAGAAATAGCGTAATGCATTGGGAGAAAAGAATGTGAGGAACCTT
GCATGGATGATGAAAATGGTTAAACAACTCGACTTCTCGCCACAGTCTGAAACTAACAAAC
AGCGGAATTAGGATTGGGGTGGGATAAACAATGAGGCGGGTCAACCGCCCGGTCTCATT
GTCTGTGCTGGTTCACTTTATCCCTCCCTCTCCCTCTGTCCAAGTGTACGATTTCTCT
AAGAATCTGGAGGTTCTGTCACCAAGTGGGACGTTCTGTGCCATGGGAATCCAGCGACTGAG
GCTGCTCGTTTCGTACCCGATCAAACCAAGGAACACTGTGTCTTTTCTCGAGCCTTCA
ATTAGGGATATTAATACTAAGCTAATGATACTCCAAGATAGCTTCAAAGATGCATTGGGA
GGAATGGTGGCCTACGCTCCAATGGATCTAAACACCGCCTGCGCTGCCATTTAGGCGAT
ATCGATCCTACCACATTCCAATCCTCCCTTCCGGTTTATGATCTCCCGTGACGGCCGT
CCTTCCGAGGGCGAAGCCGAGGGTGGCAGCTATACACTCCTCACCCTGGCTTTCCAGATC
CTTGTCTCCCGTCCSAGTTACTCTCTGATACCAACCTGGAAGTTTCTGCCACCACAGTC
AATACCTTGATTAGCTCCACCGTTCAAAGGATCAAAGCCATGCTCAAGTGCGAATGA

>G436 Amino Acid Sequence (domain in AA coordinates: 22-85)
MDFTRDDNSSDERENDVDANTNNRHEKKGYHRHTNEQIHRLETYFKBCPHPDEFQRRLLG
EELNLKPKQIKFWFQNKRTQAKSHNEKADNAALRAENIKIRRENESMEDALNNVCPPCG
GRGPGRDQLRHLQKLRAQNAAYLKDEYERVSNYLKQYGGHSMHNVEATPYLHGPNHAST
SKNRPALYGTSSNRLPEPSSIFRGPYTRGNMNTTAPPQPRKPLEMQNFQPLSQLEKIAML
EAAEKAVSEVLSLIQMDDTMWKKSSIDRLVIDPGLYEKYFTKTNTNGRPESKDVVVVQ
MDAGNLIDIFLTAEKWARLFPTIVNEAKTIHVLDSDVHRGKTFSRVIYEQLHILSPLVPP
REFMILRTCCQIEDNVWMIADV SCHLPNIEFDLSFPICTKRPSGVLIQALPHGFSKVTWI

EHVVVNDNRVRPHKLYRDLLEYGGFGYGARRWTVTLERTCERLIFSTSVPALPNNDNPGVV
QTIRGRNSVMHLGERMLRNFAMMMKMNKLDSPQSETNNSGIRIGVRINNEAGQPPGLI
VCAGSSLSLPLPPVQVYDFLKNLEVRHQWDLCHGNPATEAARFVTGSNPRNTVSFLEPS
IRDINTKLMILQDSFKDALGGMVAYAPMDLNTACAAISGDIDPTTIPILPSGFMISRDR
PSEGEAEGGSYTLTLTVAFAQILVSGPSYSPDTNLEVSATTVNTLISSTVQRIKAMLKCE*

>G556 (50..1144)

CTTTTTTGAAGCCCTTTTGACACAAAAGACCAGAACAAGTTGAAGAAATATGAATACAAC
CTCGACACATTTTGTTCACCGAGAAGGTTTGAAGTTACGAGCCTCTCAACCAAATCGG
TATGTGGGAAGAAAGTTTCAAGAACAATGGAGACATGTATACGCCTGGCTCTATCATAAT
CCCGACTAACGAAAAACCAGACAGCTTGTCTAGAGGATACTTCTCATGGGACAGAAGGAAC
TCCTCACAAAGTTTGACCAAGAGGCTTCCACATCTAGACATCCTGATAAGATACAGAGAAG
GCTAGCACAGAATCGAGAGGCAGCTAGGAAAAGTCGTTTGGCGAAGAAAGCTTATGTTCA
GCAGCTAGAGACTAGCCGGTTAAAGCTAATTCATTTAGAGCAAGAACTCGATCGTGCTAG
ACAACAGGTTTCTATGTGGGGAACGGAGTAGATACCAATGCTCTTAGTTTCTCAGATAA
CATGAGCTCAGGGATTGTTGCATTTGAGATGGAATATGGACATTGGGTGGAAGAACAGAA
CAGGCAAAATATGTGAACCTAAGAACGGTTTACATGGACAAGTTAGTGATATAGAGCTTCG
TTCTCTAGTCGAGAATGCCATGAACATTACTTTCAACTCTTCCGAATGAAGTCAGCCGC
TGCAAAAATCGATGTTTCTATGTCTATGTCGGAATGTGGAAAACCTCAGCAGAGCGGTT
TTTCTTGTGGATAGGCGGATTTAGACCCTCAGAGCTTCTCAAGGTTCTGTTACCGCATTT
TGATCCTTTGACGGATCAACAACCTTTTGGATGTATGTAATCTGAGGCAATCATGTCAACA
ATCAGAAGATGCGTTATCCCAAGGTATGGAGAACTGCAACATACATTAGCAGAGAGTGT
AGCAGCCGGGAAACTTGGTGAAGGAAGTTATATTCTCAAATGACTTGTGCTATGGAGAG
ATTGGAGGCTTTGGTCAGCTTTGTAAATCAAGCTGATCATCTGAGACATGAGACATTGCA
ACAGATGCATCGGATCTTAACCACGCGACAAGCGGCTAGAGGTTTGTAGCATTAGGGGA
GTATTTCCAAAGGCTTCGAGCTTTGAGTTGCGAGTTGGGCGGCTAGGCAACGTGAACCAAC
GTAATTAAGGTGTTTAGATGTCAAGAAAGGTTTGAACCTTAACAATCAAGAATGGAGTT
TGCTGGTGAGTGGATTTTGGGTCAAGAACAGAGCAATAACACAAGCTGCTGTGTGATG
ATGAATCTTGTCTTGGCGCTAAAGGAAATGTTTGAAGAAAGTTGTACATATGATCAGCAA
CGTAAAGTTTATAGCTTTTGAACCAACTTTTCGATGGTTGTTCTTTTTTTTTTGTAT
GTAATATTATAGATAAGCTTGTGGTATATATGATTTTAATGTGACATTACGAACCTTGATT
TATAACCATGGTAAAAAT

>G556 Amino Acid Sequence (domain in AA coordinates: 83-143)

MNTTSTHFVPPRRFEVYEPLNQIGMWEESEFKNNGDMYTPGSIIPTNEKPDLSLSEDTSHG
TEGTPHKFDQEASTSRHPDKIQRRLAQNREAAKSRRLRKAYVQQLSRLKLIHLEQEL
DRARQGFYVGNVDTNALSFSDNMSSGIVAFEMEYGHVVEEQNRQICELRTVLHGQVSD
IELRSLVENAMKHYFQLFRMKSAAAKIDVFYVMSGMWKTSAERFFLWIGGFRPSELLKVL
LPHFDPDLDQQLLDVNCNLRQSCQSEDALSQGMELQHTLAESVAAGKLGEYSYIPQMT
AMERLEALVSFVNQADHLRHETLQOMHRILTRQAARGLLALGEYFQRLRALSSSWAARQ
REPT*

>G1420 (39..1238)

AAAGTATCATCTCATAGATTCCATCTTTTCTCTATTACATGGAGAAGAAAAAGAAGAGG
ATCATCATCATCAACAACAACAACAACAAGGAGATCAAGAACACAGAGACAAAGA
TCGAGCAAGAACAAGAACAAGAACAACAAGAAATCTCTCAAGCATCATCATCAAA
ACATGGCGAATCTAGTTACGTCATCAGATCATCATCCGTTGGAGCTAGCTGGAAATCTCT
CAAGCATCTTCGATACCTCATCTTTACCTTTTCTTATTCTTATTTTGAAGATCACTCTT
CTAATAATCCTAATCTTCTCCTAGACTTGCTCCGACAGATCATCAGTTTGCTTCTTCT
CTAATTCCTCTTCTTTTTCATTGATGCCCTTCTCTCCCCAATAACAACAACAACACCT
CTTTTTTTACGGATTGCCCCTTACCTCAAGCTGAGTCATCAGAAGTCGTGAACACAACAC
CGAGTCTCCTCAAACTCAACCTCAGTCTCATCTTCTCCTCAACGAAGCTGCAAAATGATAACA
ACAGTGGTAAAGAAGTTACTGTTAAAGATCAAGAAGAGGAGATCAACAACAAGAGCAAA
AGGGTACTAAGCCACAGTTGAAGGCAAGAAGAAGAAATCAAAAGAAAGCTAGAGAAGCTA
GGTTTGGCTTTCTGACGAAGAGCGATATTGATAATCTTGACGACGGTTATAGGTGGAGAA
AATACGGCCAAAAGCTGTCAAAAACAGTCTTATCCAGAAGCTATTACCGTTGCACCA
CAGTGGGTTGCGGAGTGAAGAAGAGAGTGAGAGATCTCCGATGATCCTTCGATCGTCA
TGACAACCTACGAAGTCAAGATACCCATCCTTTCCCATGACGCCACGTGGACACATCG
GAATGCTCAGCTCAGCAATCCTAGACCAGGTGCAACCACGCGTCATCATCATCATCTT
CCATCCCTCAGCCACGTTACTTGCTGACTCAACATCACCAGCCCTACAACATGTACAACA

ACAACCTCTCTAAGTATGATCAATAGAAGATCATCCGATGGCACTTTTCGTAAATCCAGGTC
CATCATCATCATTCCTCCGGCTTTGGTTATGATATGTCTCAAGCTTCTACTTCAACTTCTT
CTTCCATTAGAGATCATGGATTGCTTCAAGATATTCTTCTTCGAGATCAGATCCGATA
CTATTAACACTCAAACCAATGAAGAGAATAAGAAATGAAGAAGTTTTTTTTTCCCGGGGCA
ATTGTTTTTTTTCTTTAGGCCGGATCCGGTAGGTAGGTTTCATGAGC
>G1420 Amino Acid Sequence (domain in AA coordinates: 221-280)
MEKKKEEDHHHQQQQQQQKEIKNTETKIEQEQEKEQKQKQISQASSSSNMANLVTSSDHHF
LELAGNLSSIFDTSSLPPPYSYFEDHSSNNPNSFLDLLRQDHQFASSNSSSFSFADFPL
PNNNNNTSFFDLPPLQAESSEVVNTTPTSPNSTSVSSSSNEAANDNNSGKEVTVKDQEE
GDQQQEKGTKPQLKAKKKKQKAREARFAFLTKSDIDNLDGGRWRKYGQKAVKNSPYP
RSYYRCTTVGCGVKRVERSSDDPSIVMTTYEGQHTHPFMTPRGHIGMLTSPILDHGAT
TASSSSFSIPQPRYLLTQHHPYNMNNNSLSMINRRSSDGTFTVNPSPSSSPGFGYDMS
QASTSTSSSIRDHGLLDILPSQIRSDTINTQTNEENKK*
>G1412 (115..1008)
CCCACGCGTCCGCCCACGCGTCCGAAACAAAAACATATAATTTGGGTTTPTAGAGTTCGA
AACTTGAATCTTTTTTTTTTTGGTTGCTGAGGAATCGAAGTAGAAGAGTATAAATGGGT
GTTAGAGAGAAAGATCCGTTAGCCAGTTGAGTTTGCCACCAGGTTTATGATTTTATCCG
ACAGATGAAGAGCTTCTTGTTCAGTATCTATGTGCGAAAGTTGCGAGGTATCATTTCTCT
CTCCAGGTCTCGGAGACATCGATCTCTACAAGTTCGATCCTTGGGATTGCGCAAGTAAG
GCTTTGTTTGGAGAGAAGGAATGGTATTTCTTTAGCCCAAGAGATCGGAAATATCCGAAC
GGGTCAAGACCAATAGAGTAGCCGGTTCGGTTATTGGAAGCAACGGGTACTGACAAA
ATTATCACGGCGGATGGTCTCGTGTGCGGATTAAAAAAGCTCTGGTCTTTACGCCGGA
AAAGCTCCCAAAGGCACTAAACCAACTGGATTATGCACGAGTATCGCTTAATAGAACAT
TCTCGTAGCCATGGAAGCTCCAAGTTGGATGATTGGGTGTTGTGTCGAATTTACAAGAAA
ACATCTGGATCTCAGAGACAAGCTGTTACTCTGTTCAAGCTTGTCGTGAAGAGCATAGC
ACGAATGGGTGCTCATCGTCTTCTTCATCACAGCTTGACGACGTTCTTGATTCTGTTCCG
GAGATAAAAGACAGTCTTTTAATCTTCTCGGATGAATTCGCTCAGGACGATTCTTAAC
GGGAACTTTGATTGGGCTAGCTTGGCAGGTCTTAATCCAATTCAGAGCTAGCTCCGACC
AATGGATTACCGAGTTACGGTGGTTACGATGCGTTTCGAGCGGCGGAAGGTGAGGCGGAG
AGTGGGCATGTGAATCGGCAGCAGAACTCGAGCGGGTTGACTCAGAGTTTCGGGTACAGC
TCGAGTGGGTTTGGTGTTCGGGTCAAACATTGAGTTTAGGCAATGAGAGAGATGTGAA
GTTACTGATGGGTGAAAAAAGTAAAAAAGAACTTGGAGATAGTAGAGTGGCAATTGATG
TAAATAATAGGGATTATATATGGGGCTTTTACCGATTGCGTGAGGCTTAGGATTCCCCAAA
GGAAAAAGGCTCGACTGGGACTAGTTTGTATCCAACCTGACGGCCCCCAATGTGTAATG
TTTCTCAACGAGAGAAAAATAATGGTTACCAATATTTTCCAAAAAAGAAAAA
>G1412 Amino Acid Sequence (domain in AA coordinates: 17-159)
MGVREKDLPLAQLSLPPGFRFYPTDEELLVQYLCKRVAGYHFLSLQVIGDIDLYKFDLPDLP
SKALFGEKEWYFFSPRDRKYPNGSRPNRVAGSGYWKATGTDKIITADGRRVGIKKALVYF
AGKAPKGTKTINWIMHEYRLIEHSRSHGSSKLDWVLCRIYKKTSGSQRAVTPVQACREE
HSTNGSSSSSSQLDDVLDSPFEIKDQSFNLPRMNSLRILNGNFDWASLAGLNPIPELA
PTNGLPSYGGYDAFRAEAGEAESGHVNRQNSSGLTQSFYSSSGFVSGQTFFBFRQ*
>G738 (1..885)
ATGGACCATCATCAGTATCATCATCATGATCAATACCAACATCAGATGATGACTAGTACT
AACAATAATTCCTATAACACCATCGTCACAACACAACCACCACCAACAACAACAATG
GATTCAACAACAGCAACAACCTATGATAATGGATGACGAGAAGAAGTTGATGACGACAATG
AGCACTAGGCCCGCAAGAACAAGAACTGTCCAAGATGCAACTCAAGCAACACCAAGTTT
TGTTATTACAACAACCTACAGCTTAGCACAGCCTAGGTACTTGTGTAAGTCTTGTGCGGAGA
TATTGGACTGAAGGTGGCTCTCTCCGTAACGTCCCCGTAGGCGGAGGTTCTAGAAAGAAC
AAGAAGCTTCCATTTCTTAATTCCTTCTTCTTCTTCCACCAAGAACCTCCCGGATCTC
AACCCTCCTTTCGTCTTCACATCATCAGCTTCATCATCAAACCTAGCAAGACGCATCAA
AACAATAATGACCTCAGCCTATCCTTCTCCTCCCCATGCAAGACAAGCGAGCTCAAGGG
CATTACGGTCAATTTCACTGAGCAAGTTGTGACAGGAGGGCAGAACTGTCTTTTCCAAGCT
CCTATGGGAATGATTCACTTTCGTCAAGAGTATGATCATGAGCACCACCAAGAAATCTT
GGGTTTTTCATTAGACAGGAACGAGGAAGAGATTGGTAATCATGATAACTTCGTTGTTAAT
GAGGAAGGAAGTAAGATGATGTATCCTTATGGAGATCATGAAGACCGTCAACAACATCAC
CATGTGAGACACGATGATGGTAATAAGAAGAGAGAAGGTGGTTCAAGCAATGAGCTATGG
AGCGGAATCATCTAGGTGGTGATAGTGGTGACCAACATGGTGA

>G738 Amino Acid Sequence (domain in aa coordinates: 351-393)
MDHHQYHHHDQYQHQMSTNNNSYNTIVTTQPPPTTTTMDSTTATTMIMDDEKKLMTM
STRPQEPNRCPRCNSSNTKFCYYNNYSLAQPRYLCKSCRRYWTEGGSLRNVPVGGGSRKN
KKLPFPNSSTSSSTKNLPDLNPPFVFTSSASSSNPSKTHQNNNDLSLSFSSPMQDKRAQG
HYGHFSEQVVTGGQNCFLQAPMGMIQFRQYDHEHPKKNLGFSLDRNEEEIGNHDFVFN
EEGSKMMPYPGDHEDRQHHHVRHDDGNKKREGGSSNELWSGIILGDSGGPTW*

>G2426 (1..1038)

ATGGGCAGATCGCCATGTTGTGATAAGGCCGGGTGAAGAAAGGCCCTTGGACTCCAGAA
GAGGATCAGAAACTTTTGGCTTATATTGAAGAACATGGCCATGGAAGCTGGCGTTCTTTG
CCTGAGAAAGCCGGTCTCCAAAGGTGTGGAAGAGTTGCAGACTCAGATGGACTAATAC
CTAAGACCTGACATCAAGAGAGGCAAATCACTGTACAAGAAAGACAAACCATCATTCAA
CTCCACGCTCTCCTCGGAAACAGGTGGTCAGCGATTGCAACTCATTTACCAAAGAGGACA
GACAACGAGATCAAGAACTACTGGAACACACACTTGAAGAAACGTCTGATCAAAATGGGG
ATAGATCCAGTGAATCACAAGCAGCAAAAACGAGACTCTTTCGTCTTCCACAGGACAATCA
AAGAACGACGACGCTTAGTCATATGGCTCAATGGGAGAGTGCAAGACTCGACGCTGAA
GCAAGGCTAGCTAGAGAATCAAAGCTTCTCCATTTACAGCATTACCAAAACAATAACAAC
CTTAACAAATCAGCAGCTCTCAACAACATTGCTTCACTCAAAAACATCAACAACTGG
ACTAAACCAAACCAAGGAAACGAGACCAACAGCTTGAATCTCCGACATCGACGGTGACA
TTCTCTGAGAATCTTCTGATGCTTTAGGAATCCCTACGGATAGCAGCAGAAATAGAAAC
AATAACAACATGAGTCTCGGCGATGATTGAATTGGCCGTATCTTCGTCAACCTCCTCC
GATGTGAGTCTGGTCAAAGAACATGAACACGACTGGATTAGGCAGATCAACTGTGGTAGT
GGAGGAATAGGAGAAGGATTACAGAGTCTATTGATCGGTGATTCGGTCCGGCCGGGTTTA
CCCACCGGAAAAACGAAGCGACGGCGGGCGTGGGAATGAGAGTGAGTATAACTACTAT
GAGGATAACAAGAATTACTGGAATAGCATTCTCAACTTGGTTGATTCTTCACCGTCCGAT
TCCGCGACGATGTTCTGA

>G2426 Amino Acid Sequence (conserved domain in AA coordinates:14-114)

MGRSPCCDKAGLKKGPWTPEEDQKLLAYIEEHGHSWRSLEKAGLQRCGKSCRLRWNTNY
LRPDIKRGKFTVQEEQTIIQLHALLGNRWSAIATHLPKRTDNEIKNYWNTHLKKRLIKMG
IDPVTHKHKNETLSSSTGQSKNAATLSHMAQWESARLDAEARLARESCLLHLQHYQNNNN
LNKSAAPQQHCFQKTSTNWTNPKPNQNGDQQLSPTSTVTFSENLLMPLGIPTDSSNRN
NNNNESSAMIELAVSSSTSDVSLVKEHEHDWIRQINCGSGGIGEGFTSLIGDSVGRGL
PTGKNEATAGVNESEYNYEDNKNYWNLSILNLVDSSPSDSATMF*

>G1524 (1..825)

ATGGGGAGAACTAAGGAGCAGGCAACATTAAGTCCGATATCCACCCTGTCTTAGGAATCCT
GCTAAATTCAATGATATAAACAAGCACTCCAGGAAAAAGGATATGGTAAGGCTCTGAAA
AGAAAAACCTTGGACGGGTGTGACATGCCCTGTCTGTCTTGAGGTTCTCACAACCTCGGTC
GTCTCTCTTTGTTTCATCTTACCACAAAGGATGCCGTCCGTACATGTGTGCCACGGGAAAC
CGTTTCTCAAATTGTCTAGAGCAGTACAAAAAGGCATATGCCAAGGATGAGAAAAGTGAC
AAACCGCCAGAGCTATTGTGCCCCGCTTTGTAGGGGTGAGGTGAAAGGCTGGACCGTTGTG
GAAAAGGAACGTAAGTATCTGAATTCTAAGAAAAGGTGATGCTGAACGACGAGTGTGTTG
TTTTATGGAAGCTATAGACAGCTCAAGAAGCATGTTAAGGAGAACCATCCGAGAGCCAAG
CCAAGAGCCATAGACCTGTGTGTTGAGGCGAAATGGAAGAAGCTTGAGGTTGAGAGGGAG
AGGAGTGATGTAATCAGCACAGTGTGTCGTCACACCTGGGGCTATGGTATTTGGAGAC
TATGTGATTGAGCCATACAATGGTTATGATCATCAAGATGACAGTGACGATTACAGTGAT
TCGTCCGATGACGAAATGGAAGGTGGGGTATTTCGAGCTTGGAGCATTCGACCTGGGCCGT
CTTCAACCGCGTTCCGCTGCCATCTCAAGCCGGGGAATTCCGCGTATGATCATAAGGAAC
CGGTGGGCTCGAAGCAGAGGTGCGAGCAGAAGGCGACAAACATAA

>G1524 Amino Acid Sequence (conserved domain in AA coordinates:49-110)

MGRTEQATLTRYPPCPNPAKFNDINKALQEKYGKALKRPWTVTCPCVLEVPHNSV
VLLCSSYHKGCRPYMCATGNRFSNCLEQYKAYAKDEKSDKPPPELLCPLCRGQVKGWTVV
EKERKYLNSKKRSCMNDECLFYGSYRLKHHVKENHPRAKPRADPVLEAKWKKLEVERE
RSDVISTVMSSTPGAMVFGDYVIEPYNGYDHQDDSDSDSDEMEGGVFELGAFDLGR
LQPRSAAISSRGIRMIIRNRWARSRGASRRRQT*

>G1243 (1..3174)

ATGGCGAGAAATTCGAATTCGATGAGGCTTTCTCGTCAGAGGAGGAAGAAGAGCGGGTT
AAGGATAATGAAGAAGAAGATGAGGAGGAGCTCGAGGCTGTTGCTCTTCTGCTCC
GACGATGACGAAGTAGCCGCCCGCCGACGAATCACCAGTCTCCGACGAGAGGCTGCTCCC

GTAGAAGATGATTACGAGGACGAAGAAGATGAGGAAAAAGCTGAAATCAGCAAACGTGAG
AAAGCCAGACTTAAAGAGATGCAGAAGTTGAAGAAGCAGAAGATTCAAGAGATGCTGGAG
TCGCAGAATGCTTCCATTGACGCGGATATGAACAATAAGGGAAAAGGGAGACTGAAGTAT
CTTCTGCAGCAAACCTGAGTTATTTGCCCACTTTGCTAAAAGTGATGGATCTTCTTCTCAG
AAGAAGGCAAAAGGAAGGGGACGTCATGCTTCCAAAATAACTGAAGAGGAGGAAGACGAA
GAGTATCTAAAGGAAGAAGAGGATGGCTTAACTGGATCTGGAAAACACACGGTTACTCACA
CAGCCCTCTTGTATTCAAGGGAAGATGAGAGATTACCAATTAGCTGGTTTGAAGTGGCTC
ATTCTGCTTTTATGAGAATGGCATAAATGGAATCTTGTGCTGATGAAATGGGTCTGGGGAAG
ACGCTTCAAACGATTTCTTTGTTGGCATATCTTCATGAATACAGGGGAATCAATGGTCCC
CATATGGTGGTTGCTCCAAAATCAACACTTGGTAATTGGATGAACGAAATTCGCCGGTTT
TGCTCTGTCTACGTGCTGTGAAGTTCTTGGTAATCCTGAGGAGAGGAGACATATTCGA
GAAGACCTGCTAGTTGCTGGGAAATTTGATATTTGTGTCAAGCTTTGAGATGGCCATC
AAAGAGAAGACAGCACTTCGTCGGTTTAGCTGGCGTTATATTATCATTGATGAAGCGCAT
CGAATCAAGAACGAGAATTCAC'TCCTTTCTAAAACCATGAGACTTTTTAGCACCAATTAT
CGGCTTCTTATCACGGGGACCCCCCTTCAGAATAATCTCCATGAAGTGTGGCTCTTCTA
AATTTTCTTCTGCCTGAGATTTTGTAGTTCAGCAGAGACTTTTGATGAATGGTTTCAAATT
TCTGGTGAGAATGACCAGCAAGAAGTTGTGCAACAACCTGCACAAGGTTCTTCGACCATT
CTTCTTCGAAGACTAAAGTCAGATGTTGAGAAAGGTTTGCACCGAAGAAGGAGACCATA
CTTAAAGTTGGTATGTCTCAGATGCAAAAGCAATACTACAAGGCTTTACTGCAGAAGGAT
CTTGAAGCGGTTAATGCTGGTGGAGAACGCAACGCTCTGCTAAACATTGCAATGCAACTG
CGTAAATGCTGCAATCACCCCTATCTCTTCCAGGGTGCAGAACCTGGTCCCCCATATACC
ACAGGAGATCACCTTATAACAAATGCTGGTAAGATGGTTCTCTTGGATAAATTGCTTCCT
AAGTTGAAAGAACGTGATTCAAGGGTGTGATATTTTCTCAGATGACAAGACTTTTGGAT
ATTCTTGGAGACTATTTAATGTATCGTGGTTACTTGTATTGCCGTATTGATGGAACACT
GGTGGTGACGAACGAGATGCCTCCATAGAAGCCTACAACAAGCCAGGAAGTGAGAAATTT
GTTTTCTTGTATCTACTAGAGCTGGAGGGCTTGGTATCAATCTTGTCTACTGCAGATGTT
GTGATCCTTTACGATAGTGATTGGAACCCACAAGTCGACTTGCAAGCTCAGGATCGTGCC
CATAGGATTGGTCAAAAAAAGAAGTTCAAGTGTTCGATTCTGCACTGAGTCTGCTATT
GAGGAGAAAGTGATTGAAAGAGCTTACAGAAGTTAGCACTTGATGCTCTGGTTATTCAA
CAAGGGAGATTGGCAGAACAGAAAAGTAAGTCTGTCAATAAGGATGAGTTGCTTCAAATG
GTAAGATATGGTGTCTGAGATGGTGTTCAGTTCTAAAGATAGCACAAATCACAGACGAGGAT
ATTGATAGAATCATTGCCAAAGGAGAAGAGGCAACAGCTGAAC'TTGATGCTAAGATGAAG
AAATTCACAGAAGATGCTATACAGTTTAAAATGGATGACAGTGTGACTTCTATGATTTT
GATGATGACAATAAGGATGAAAACAAGCTCGATTTTAAAAGATTGTAAGCGACAATTGG
AATGATCCCCCAAGCGGGAGAGAAAAGCGCAACTACTCTGAATCTGAGTACTTTAAGCAA
ACATTGCGGCAAGGTGCTCCAGCTAAACCTAAAGAGCCTAGAATTCGCGCATGCCCCAG
TTGCACGATTTCCAGTTCTTTAACATTAGAGATTGACCGAGTTGTATGAAAAGGAAGTA
CGTTATCTCATGCAAAACATCAGAAAAATCAGTTGAAAGACACAATTGATGTTGAAGAA
CCAGAAGGTGGGGATCCCTTAAC'TACTGAAGAAGTAGAAGAAAAGGAGGATTATTGGAG
GAGGGTTTCTCAACATGGAGCAGAAGAGATTTTAACTTTCTCAGGGCTTGTGAGAAG
TATGGCCGCAACGACATAAAAAGCATTGCCTCTGAGATGGAAGGGAAAACAGAGGAAGAA
GTTGAAAGATATGCCAAAGTATTTAAAGAGCGGTACAAGGAGCTGAACGACTATGATAGA
ATCATTAGAACATTGAGAGGGGAGAGGCAAGGATCTCTAGGAAAGACGAAATCATGAAG
GCCATAGGGAAGAAACTGGATCGCTACAGAAACCCTTGGCTGGAACTGAAGATTCAATAT
GGTCAGAACAAAGGCAAGCTGTACAATGAAGAGTGTGACCGTTTCATGATCTGCATGATT
CACAAACTTGGTTATGGGAATTGGGATGAGCTAAAGGCAGCATTTAGGACATCGTCTGTG
TTCAGGTTTGTACTGGTTTGTGAAATCCCGCACGAGTCAGGAACCTTGAAGAGATGCGAC
ACTCTGATTGACTGATCGAGAAAGAGAACAGGAGTTTGTATGAAAGAGAGAGGCAAGCC
CGCAAAGAGAAGAAGCTCGCGAAGAGTGCAACACCATCAAAGCGACCTTTAGGAAGACAA
GCAAGTGAGAGTCTTTCATCGACGAAGAAGCGGAAGCACCTGTCTGATGAGATGA
>G1243 Amino Acid Sequence (domain in AA coordinates: 216-609)
MARNSNSDEAFSSEEEERVKNDEEEDEEELEAVARSSGSDDEVAADSPVSDGEAAP
VEDDYEDDEDEEKAEISKREKARLKEMQKLKQKIQEMLESQNASIDADMNNKGKGRLLKY
LLQOTELFAHFASDSSSQKKAKGRGRHASKITEEEDEEYLKEEDGLTSGNTRLLT
QPSCIQGMKRDYQLAGLNLIRLYENGINGILADEMGLGKTLQOTISLLAYLHEYRGINGP
HMOVAPKSTLGNWMNEIRRFPCVLRVAKFLGNPEERRHIREDLLVAGKFDICVTSFEMAI
KEKTALRRFSWRYIIIDEAHRICKENSLLSKTMRLFSTNYRLLITGTPLQNNLHELWALL

[illegible]

MDKEKSPAPPSPGGLPPSPGRYSAFSPNGSSFAMKAESSFPPLTPSGSNSSDANRPSHDI
SRMPDNPPKNLGHRRAHSEILTLPPDDLSDLDLGVGAADGPSFSDDTDEDLLMYLDME
KFNSSATSTSQMGEPSEPTWRNELASTSNLQSTPGSSSERPRIHQHSQSMGSTTIKPE
MLMSGNEDVSGVDSKKAISAAKLSELALIDPKAKRIWANRQSAARSKERKMRYIAELER
KVQTLQTEATSLSAQLTLLQRDTNGLGVENNELKLRVQTMEQQVHLQDALNDALKEEVQH
LKVTLGQGPSNGTSMNYGSFGSNQQFYPNNQSMHTILAAQQLQQLQIQSQKQQQQQQHQ
QQQQQQQQQFHFQQQQQLYQLQQQRLQQEQQSGASELRRPMPSPGQKESVTSPPDRETPL
TKD*

ATCGGTTGGATCGATGGCGGAGAGAGCAAGGCAGGCCAACATTCTCCACTAGCGGGACCC
CTAAAGTGTCCTCGATGCGACTCCAGCAACACTAAGTTCTGTTACTACAACACTATAAC
CTCACTCAGCCTCGTCACTTCTGCAAAGGTTGCCGTGCTACTGGACACAAGGGGGCGCC
CTGAGAAACGTCCTCTGTAGTGGAGCGCTGCCGGAGGAATAACAAGAAGGGCAAAAATGGA
AATTCAAATCTTCTTCTTCTGCTCCAAACAGTCTTCTCGGTCAACGCTCAAGTCTCT
AGCTCAGGACGACTAAGACAAATCATCAGTTCCTCTTTTCAACCAACTCTTTTCAACATCTC

ACTCAACTCGGAGGTATTGGTTTGAAGCTTAGCCGCTACTAATGGCAACAACCAAGCTCAC
CAGATCGGTTCCAGTTTGATGATGAGCGATCTAGGGTTTCTCCATGGACGAAATACTTCA
ACTCCGATGACGGGAAACATTCATGAAAACAACAATAATAACAATGAAAACAACCTA
ATGGCATCCGTTGGATCTTTGAGCCCTTTGCTCTCTTCGATCCAACGACGGGGCTATAC
GCTTTCAGAACGACGGTAATATCGGGAACAACGTTGGGATATCTGGTTCTTCTACTTCC
ATGGTTGATTCTAGGGTTTATCAGACGCCTCCGGTGAAGATGGAAGAACAACCTAATTTG
GCTAACTTGTCTAGACCGGTCTCCGGTTGACGTCCTCGGGAATCAAACAAATCAGTAC
TTTTGGCCTGGTTCGGATTTCTCGGGTCCTTCTAATGATCTCTTGTGA

>G1909 Amino Acid Sequence (conserved domain in AA coordinates:23-51)

MGGMAERARQANIPPLAGPLKPCRDSSNTKFCYNNYNLTQPRHFCKGCRRYWTQGA
LRNVPVGGGCRNNKKGKNGNLKSSSSSSKQSSSVNAQSPSSGQLRTNHQFPFSPPLYNL
TQLGGIGLNLAAATNGNNQAHQIGSSLMSDLGLFHLGRNTSTPMTGNIHENNNNNNNNNL
MASVGSLSPPFALFDPPTGLYAFQNDGNIGNNVGISGSSSTSMVDSRVYQTPPVKMEEQPNL
ANLSRPVSGLTSPGNQTNQYFWPGSDFSGPSNDLL*

>G1663 (64..630)

TTCTCTCTGTGAATCCTTGTTCATCGTCACTGAAATTAGTTTACAAAATCGACGAATTTCG
GAGATGATTTTTTCAGAAATGTGTGCAGAAATGAGTCCAACCTTCAACGCTATAGCTTCCGAA
TCGCGTTCCCAAACGCAGTTCGGTGTTCGAAATCCTCCTCGACGGCGGGCGGATGTATC
TCCGCCAGGACTAAAGACCGTCACACGAAGGTTAACGGACGAAGCCGTCGAGTTACGATG
CCGGCTCTCGCCGCCGCTAGGATTTTCCAGTTAACCGGTGAGCTCGGTACAAAACCTGAA
GGAGAAACCATCGAATGGCTTCTTAGTCAAGCTGAACCGTCGATTATTGCCGCCACTGGC
TACGGGACTAAGCTCATTTTCAATTTGGGTTGATGTTGCGGCGGACGATTCTCGTCTGTCG
TCGTGATGACGTGCGCGCAACGCAACGCAACGCCACAATCGCCGAGTTGTAGGTTG
GATCTTTGTGACGCAATCGGAATTCAGTATCCGGTGAATGGTTACAGTCATATGCCGTTT
ACAGCGATGCTTTTAGAGCCGATGACCACGACGCGGAATCTGAGGTTGAGATCGCGGAG
GAGGAGGAACGTAGACGCCGTCACCATTAGTAAAATTAGGCTTTTGATTTAGAGTGTTAA
AATTAGGATTTTAAAGTTTAGGAGGTAACAGATAAGGATAATT

>G1663 Amino Acid Sequence (domain in AA coordinates: TBD)

MIFQNVCRNESNFNAIASESRSTQFGVSKSSSSGGGCISARTKDRHTKVNGRSRRVTMP
ALAAARIFQLTRELGHKTEGETIEWLLSQAEPSSIIAATGYGTKLISNWVDVAADSSSSS
SMTSPQTQTQTPQSPSCRLDLCPIGIQYPVNGYSHMPFTAMLLEPMTTAESEVEIAEE
EERRRRHH*

>G1231 (103..870)

CAAACCCAAATCTCTCAGCGCCGGTCAAATACTTGTCTCTCTCTCTCTCTCTTTTAC
TCTTGTCTTGTCTCTCTTGAAGCTGTTTGTCTGTAAAGATGGAAGCAGGTGGCGCG
TACAATCCACGCACTGTTGAAGAGGTGTTTAGGGATTTTAAAGGGTCGTAGAGCTGGCATG
ATTAAGGCTTTAACCAGTATGTTTCAAGAGTTTTCGACTTTGTGATCCCCGAAAAGGAG
AACCTTTGCCTTTACGACATCCAAATGAGCACTGGGAAGTGAATTTGCCAGCTGAAGAG
GTTCTCTCTGAGCTCCCAGAGCCTGTCTTGGGTATCAATTTTGCCAGAGACGGGATGGCG
GAAAAGGATTGGTTGTCCCTTGTGCTGTCCACAGTGATGCTTGGCTTCTTGTCTGTGCT
TTCTTTTTTGGAGCCAGGTTTGGATTTGACAAAGCTGATAGGAAGAGGCTTTTCAATATG
GTGAATGACCTCCCAACAATCTTTGAGGTTGTAGCTGGCACTGCTAAGAAACAAGGAAAA
GATAAGTCTCTGTTTCCAACAACAGCAGCAACAGATCCAAATCAAGCTCCAAGCGAGGA
TCTGAATCCCGTGCCCAAGTTCTCAAAGCCGAGCCCAAAGATGATGAGGAGGAGGAAGAG
GAAGGTGTGGAAGAGGAGGATGAGGATGAGCAAGGTGAACACAGTGTGAGCATGTGGT
GAGAGCTATGCAGCTGATGAGTTCTGGATTGCTGTGACCTCTGTGAGATGTGGTTTCAT
GGAAAGTGTGTTAAGATAACACCAGCAAGAGCTGAGCACATCAAGCAATACAAGTGCCCT
TCTTGCAGCAACAAAAGGCTCGTTTCTTAAATTTGTTGACCGCTCGCTTCTGTGTATCTA
CCTTTGCATATGATGATGAACAGCTTAACTGTTTGGTTTAGATCAGATTTGTTCATATGA
TTTGGTAATTTAGGAAGACATTTTAGTTTTCATTGTTACATTTTGGCGATTGAAGGGA
TAACTCTTTGTTTAGGGTAATGATCTTTGCTCTGTTTATGTTTATTAAACATTC
TTCAAACCTCAATCAAAGTATTTTGGTTAGTCTTAAAA

>G1231 Amino Acid Sequence (domain in AA coordinates: TBD)

MEAGGAYNPRTVEEVFRDFKGRRAGMIKALTVDVQEFRLCDPEKENLCLYGHNPNEHVEV
NLPAAEEVPELPEPVLGINFARDGMAEKDWLSLVAVHSDAWLLAVAFFFGARFGFDKADR
KRLFNMVNDLPTIFEVVGATAKKQKDKSSVSNNSSNRKSSSKRGSESRKFSKPEPKD
DEEBEEBVEEEDDEQGETQCGACGESYAADEFWICDLCCEMWFHGKCVKITPARAEHI

>G227 (21..983)

>G227 Amino Acid Sequence (domain in AA coordinates: 13-112)
MSNPTRKNMERIKGPWSPEEDDLLQRLVQKHGPRNWSLISKSIIPGRSGKSCRLRWCNQLS
PEVEHRAFSQEEDETIIRAHARFGNKWATISRLLNGRTDNAIKNHNWNTLKRKCSVEGQS
CDFGGNGGYDGNLGEEQPLKKTASGGGGVSTGLYMSPGSPSGSDVSEQSSGGAHVFKPTV
RSEVTASSSGEDPPTYLSLSLPWTDETVRVNEPVQLNQNTVMDGGYTAELEFPVRKEEQVE
VEEEEAKGISGGFGFMTVQEMIRTEVRSYMDLRNVGGSSSGGGGGSCMPQSVN
SRRVGFREFIVNOIGIKME*

ACTATTACATGCTCTTCTCGCTTCAA^{AA}ACGGCACCGTTTCCACTTGTTATTATTTTCTCTCTATCGTCTAA^{CA}AAAAAAAAAACTGACTTGGGATTTTTTTTTCATTGTCTAGCCCAAAGAAGAAGATAGAAACGAAGAAAAAAGCAAACACATTTTGGGTCCCCGGTGTTAGGATCAAATTAGGGCACAACCTTATCGGAGAAAGAAGCCATGGGAAGAAGAAAAGTCGAGATCAAGCGAATCGAGAA^{CA}AAAAGCAGTCGACAAGTCACTTTCTCCAAACGACGCAAAGGTC TCATCGAAAAGCTCGACAAC^{TT}CAATTCTCTGTGAATCTTCCATCGCTGTGTGTCGCCCTCTCCGGTTCGGGAAA^{CT}CTACGACTCTGCCTCCGGTGACAACATGTCAAAGATCAATTGATCGTTATGA^{AA}TATCATCATGCTGATGA^{ACT}TAAAGCTTAGATCTTGCAGAAAAAATTCGGAATTATCTTCCACACAAGGAGTTACTAGAAATAGTCAAAGCAGCTTGAAGAATCAATGTGCGATAATGTAAGTGTAGATTCTCTAATATCTATGGAGGAACAGCTCGAGACTGCTCTGTCTAGTAATTAGAGCTAAGAAGACAGAACTAATGATGGAGGATATGAAGTCACTTCAAGAAAGGGAGAAGITGCTGATAGAAGAGAACCAGATTCTGGCTAGCCAGGTGGGGAAGAAGACGTTTCTGGTTATAGAAGGTGACAGAGGAATGTACGGGAAAATGGCTCCGGGCAACAAAGTACCGGAGACTCTTTCGCTGCTCAAGTAATCACCATCATCAACGGCTGAGCTTTCACCATAACTTACTCACAGCCTGATT^{CA}GAAAGCTTTTACAAAATTGTAAATTATAAAAAGCTGCA TAATAATCTCAACCTTTTTATCTTCTCTCGCCAAATGTGGAATAAAGGTAAAAACAAAACGAAGCTCTTTTCTTTATGCGAAAGCTTGTAAACTAGGATAAAGTACCGATCTTTGTGTACTCTTAGTAGACAAATATCAGAGTTC^{TT}TGTGCTGTG

>G1842 Amino Acid Sequence (domain in AA coordinates: 2-57)
MGRRKVEIKRIENKSSRQVTFSKRRKGLIEKARQLSILCESSIAVVAVSGSGKLYDSASG
DNMSKIIDRYEIIHHADELKALDLAEKIRNYLPHKELLEIVQSKLEESNVDNVSVDSLISM
EEQLETALSVIRAKKTELMMEDMKSLOEREKLLIEENQILASQVGKKTFLVIEGDRGMSR
ENGSGNSQVPETLSLK*

ATGGATGATATAGCGGAACCTTGAATGGTTATCAAATTTCTGTAGATGATTCTTCTTTCACG
CCGTATTCTGCTCCGACGAATAAACCGGTTTGGTTAACCGGAAATCGGAGACATCTGTAT
CAACCGGTTAAAGAGGAGACCTGTTCAAATCCCAACATCCGGCCGTCAAACCCAGACCC
AAACGAGCCAGAACC GGAGTCAGAGTCTGGTCTCATGGTTCGCAGTCGTTAAACCGCATCA
TCTTCAAGCTCTACAACATCTTCGTCGTCCTCTCTCTCTCAAGCCCTCTATGCGTCT
GCCACGGTCAGTTTCTTGTATGAGCCAATGACTAAACACAAAGAAGAAGAAAGTTTGG

AAAAACGCTGGTCAGACGCAAACGCAAACGAGACGAGACGCGGCAGTGTGGTCATTGT
GGAGTTCAGAAAACGCCGAGTGGAGAGCAGGACCATTAGGAGCGAAGACGTTGTGTAAT
GCGTGTGGTGTGCGTTACAAATCGGGTTCGGTTACTACCCGAATATAGACCCGCTTGTAGC
CCAACATTTTCGAGTGAGCTTCACTCAAACCACCACAGTAAAGTCATTGAGATGCGTAGG
AAGAAAGAGACTTCTGACGGTGTGTAAGAAACCGGTTTGAACCAGCCGGTTCAGACGGTT
CAGGTTGTCTCGAGTTTTTGA

>G1505 Amino Acid Sequence (domain in AA coordinates: TBD)
MDDIAELEWLSNFVDDSSFTPYSAPTNKPVWLTGNRRHLVQPVKEETCFKSQHPAVKTRP
KRARTGVRVWSHGSQLTDSSSSSTSSSSSPRPSPLWLASGQFLDEPMTKTQKKKKVW
KNAGQTQTQTQTQRCGHGCVQKTPQWRAGPLGAKTLCNACGVRYKSGRLLPEYRPACS
PTFSSSELHNSNHKSVIEMRRKKETSDGAEETGLNQPVQTVQVVSSF*

>G657 (1..2331)

ATGAAGCGTGAGATGAAAGCACCTACTACTCCACTAGAGAGTCTCCAAGGTGACCTCAAA
GGAAAACAAGGGAGGACATCTGGCCCTGCTAGACGATCTACCAAAGGACAATGGACACCT
GAAGAGGACGAAGTCTTGTGTAAAGCTGTTGAGCGTTTTCAAGGAAAGAACTGGAAGAAG
ATAGCTGAATGTTTTTAAGGATCGGACTGATGTTCAAGTGTCTTATAGATGGCAAAAGGTC
TTGAACCCAGAGCTTTGTGAAAGGACCGTGGTCAAAAGAGGAGGATAACACAATAATTGAC
CTGGTTGAAAAATATGGGCCAAAGAAATGGTCTACTATATCTCAGCATTTACCTGGGCGC
ATAGGAAAGCAATGTAGGGAAGGTGGCATAACCATCTTAACCTGGGATTAATAAAAAAT
GCATGGACTCAGGAAGAGGAACTGACTCTTATTCGTGCGCATCAAAATTATGGGAATAAA
TGGGCAGAGCTTATGAAATTTTTGCCAGGAAGGTGAGATAATTTCGATAAAAAATCATTGG
AACAGCTCAGTTAAGAAGAAGTTGGATTCTACTATGCATCAGGTCTTTTAGATCAGTGT
CAAAGCTCGCCATTAAATGCGCTTCAGAACAATCTATCGCTTCATCTTCTCGTGGATG
CACAGCAATGGAGATGAAGGTAGTTCAAGGCCAGGGGTTGATGCTGAGGAATCAGAATGC
AGCCAAGCTTCAACTGTTTTCTCACAATCAACCAACGATTTACAAGATGAAGTTCAACGT
GGAAATGAGGAATATTACATGCCTGAATTTCAATCAGGAACGGAGCAGCAATCTCAAC
GCTGCATCTCATGCAGAACCGTACTACCTTCTTTAAAGATGTCAAATTGTTGTCCCC
GAAATTTCTTGTGAAACAGAATGTTTCAAGAAGTTTCAAGAATCTTAATTGTTCTCAGAG
CTAAGAACTACCACAGCTACGGAGGATCAATTGCCGGGTGTATCTAATGATGTAAACAG
GACCGTGGTCTAGAGTTATTGACCCATAACATGGACAACGGTGGAAAAACCAAGCACTT
CAACAAGATTTTTCAAGTTTCAGTAAGATTAAAGTATCAACCTTTTTTGTCAAACCTCGGAC
ACAGATCCAGAAGCTCAAACCTTTGATCACGGATGAGGAGTGTGTAGGGTTCTTTTCCA
GATAACATGAAAGATAGCAGTACATCTTCTGGTGAGCAAGGTGCGAATATGGTTGACCT
CAAAACGGCAAAGGATCTCTTTGTTCTCAGGCTGCAGAAACCCATGCTCATGAACTGGA
AAAGTTCCAGCTTTACCGTGGCATCTTCAAGTTCTGAGGGCCTGGCGGGTCATAATTGT
GTCCCTTTGTTGGATTACAGCTTGAAGGACTCACTTTTACCCCGTAATGATTTCAACGCT
CCTATACAAGGTTGTCGCTTTTTGAGCTACCGAATTAGAATGTAAGACTGATACAAAT
GACGGTTTCATCGATACTTACGGACATGTAACCTCCCATGGCAATGATGATAATGGTGGT
TTCCAGAAACAACAGGGGCTGTATATATTTCCCAAGGATTTCTTGAAGCTAGTACCTTTG
AATAGTTTTTCTTCTCTCTTAGAGTGAACAAGATTTATTTTCTATTGACGATAAGCCG
GCTGAAAAAGACAAAGGAGCTCTTTGTTATGAACCTCCACGTTTCCAAGTGCAGATATT
CCTTTCTTCAGCTGTGATCTTGTACCATCAAATAGTGACTTACCGCAAGAGTACAGTCCC
TTTGGTATCCGTGAGTTGATGATTTCTTCAATGAATTGTACAACCTCCGTTAAGGTTATGG
GATTCACCGTGTACGATAGGAGCCCTGATGTCATGCTTAATGATACTGCCAAAAGTTTTT
AGTGGTGCACCATCCATCTTAAAGAAGCGGCATCGAGACTTGCTTTCACCTGTGCTTGAT
AGAAGAAAAGACAAAAGCTTAAAGGGCTGCGACTTCTCTCTTGGCTAATGATTTTTTCG
CGCTTAGATGTAATGCTTGATGAAGGAGATGATGTCATGACCTCTCGTCCGTCAGAGTCT
CCTGAAGATAAAAAATATATGTGCTCCCTTCCATAGCCAGAGATAACAGAAATTGTGCA
TCAGCTCGGTTATATCAAGAAATGATTCCGATAGATGAGGAACCAAGGAAACCTTAGAA
TCAGGTGGAGTGACTTCTATGCAAAATGAAATGGATGTAATGACGGTGGTGTCTCAGCT
AAAAATGTAAGTCCGCTTTGTCTTGCATATTATCTGGTATCAGTTATAA

>G657 Amino Acid Sequence (domain in AA coordinates: TBD)
MKREMKAPTPPLESLQGLDKGKQGRSTSGPARRSTKGQWTPPEDEVLCKAVERFQGNWKK
IAECFKDRITDVQCLHRWQKVLNPELVKGPWSKEEDNTIIDLVEKYGPKKWSTISQHLPGR
IGKQCRERWHNHLNPGINKNAWTQEEELTLIRAHQIYGNKWAELMKFLPGRSDNSIKNH
NSSVKKKLDSEYASGLLDQCQSSPLIALQNKSIASSSSWMHNSNGDEGSSRPVDAEES
SQASTVFSQSTNDLQDEVQRGNEEYMPFHSQTEQQISNAASHAEPYPYPSFKDVKIVVP

EISCETECSKKFQNLNCSHELRTTTATEDQLPGVSNDKQDRGLELLTHNMDNGGKNQAL
QQDFQSSVRLSDQPFLLNSDTPDPAQTLITDEECCRVLPDNDMKDSSTSSGEQGRNMVDP
QNGKGSLSQAAETHAHETGKVPALPWHPPSSSEGLAGHNCVPLLDSDLKDSLLPRNDSNA
PIQGCRFLFGATELECKTDTNDGFIDTYGHVTSNGNDNGGFPEQQGLSYIPKDSLKLVLPL
NSFSSPSRVNKIYFPIDDKPAEKDKGALCYEPPRFPADIPFFSCDLVPSNSDLRQEYSP
FGIRQLMISSMNCCTPLRLWDSFCHDRSPDVMLNDTAKSFSGAPSILKKRHRDLLSPVLD
RRKDKKLKRAATSSLANDFSRLDVMLDEGDDCMTSRPSESPEDKNICASPSIARDNRNCA
SARLYQEMIPIDEEPKETLESQGVVTSMQNENGENDGGASAKNVSPSLSLHIIWYQL*

>G1959 (141..1028)

CGTCGACTGTCCATAAATCCGGAGCCTGACCCGACGTTTGACCCGGATCCGAAACTCCCA
CAATCTCCATACCACCCAAATTCATCTCCCTAAAGCTTCTCTCATTTCCTCCGGGAAAA
TCGGCGACCAAAATTGAAAAATGTACTCAGCGATTTCGTCGCTTCCACTCGATGGTGGAC
ACGTTGGTGGTGACTACCATGGACCTCTTGACGGAACCAATCTTCCGGTGACGCTTGT
TGTTTTAAAGACTGACCTTAAACCTCGTCTCCGGTGGAACACTGAGCTTCATGAGAGAT
TCGTTGACGCCGCTTACTCAGCTCGGTGGTCTGACAAAGCGACTCCCAAATATTATGA
GAACAATGGGAGTGAAGGGTCTCACTCTCTACCACTCAAATCACATCTTCAGAAATCC
GCCTAGGGAGGCAAGCTGGCAAGAATCAACTGAGAACTCTAAAGATGCTTCTGTGTAG
GGGAGAGTCAGGACACAGGTTCACTTCGACATCATCAATGAGAATGGCGCAGCAGGAGC
AGAACGAGGGTTACCAAGTCACCGAAGCTCTACGTGCTCAGATGGAAGTCCAAAGAAGAC
TACACGATCAATTGGAGGTGCAACGGAGGCTCCAGCTGAGGATAGAGGCACAAGGAAAT
ACCTGCAATCGATTCTTGAAAAAGCTTGCAAGGCCTTTGACGAGCAAGCTGCTACTTTTG
CTGGACTTGAGGCTGCTAGGGAAGAGCTATCAGAGCTAGCCATCAAAGTCTCCAATAGCT
CTCAAGGAACATCAGTCCCGTACTTCGATGCAACAAAGATGATGATGATGCCATCGTTGT
CAGAGCTTGCACTAGCAATAGACAACAAAAACAACATCACAACTGTTCACTAGAAA
GCTCTCTGACTTCCATCACACATGGGAGCTCTATATCTGCTGCATCAATGAAGAAGCGTC
AACGTGGAGACAATTTGGGCGTAGGGTATGAATCAGGCTGGATTATGCCTAGTAGACCA
TTGGATAAAGTTTAGGAGAGGGAAAAAGTTCATTATGGGAAAGGTAGAGATAAGATTTAA
CTGTTCTTTACTTGTCTTGAGGGCCTGCGGCCGCT

>G1959 Amino Acid Sequence (conserved domain in AA coordinates:46-97)

MYSAIRSLPLDGGHVGGDYHGPLDGTNLPDACLVLTTDPKRLRWTTTELHERFVDAVTQ
LGGPDKATPKTIMRTMGVKGLTLYHLKSHLQKFRLLGRQAGKESTENSKDASCVGESQDTG
SSSTSSMRMAQQEQNEGYQVTEALRAQMEVQRRLHDQLEVQRRLQLRIEAQGYLQSI
KACKAFDEQAATFAGLEAAREELSELAIKVSNSSQGTSPYFDTAKMMMPSLSELAVAI
DNKNNTTNCSSVESSLTSITHGSSISAASMKKRQRGDNLVGVYESGWIMPSSTIG*

>G2180 (1..1440)

ATGGCTCCTGTCTCGTTACCTCCAGGTTTCCGATTCCATCCAACAGACGAGGAACTAATT
ACTTACTATCTAAAAAGAAAGATCAACGGTCTAGAAATCGAACTTGAAGTTATCGCTGAA
GTTGATCTTTACAAGTGTGAGCCATGGGACTTACCAGGGAAGTCTTGCTTCCGAGCAAA
GACCAAGAATGGTACTTCTTCAGCCACGAGACCGGAAGTATCCCAACGGCTCAAGGACA
AACC GGGAAC TAAAGGCGGTTATTGGAAGGCTACAGGTAAGACCGCCGAGTTAGTTGG
AGAGACCGAGCCATAGGAACCAAGAAGACATTGGTTTACTACCGTGGGCGCGGCCACAT
GGCATAAGAACTGGTTGGGTCTGTCACGAATATCGACTTGATGAAACAGAATGTGAGCCT
TCTGCATACGGCATGCGAGACGCATATGCACTTTGTGCTGTGTTCAAAAAGATTGTTATT
GAAGCTAAGCCAAGAGATCAACATCGGTCAATGTCCACGCGATGTGCAATGTGAGTGGT
AATTGCTCATCGAGTTTGTGACTTGTTCGGATCTCGAAATCAGTTCAACTACTCATCAA
GTTCAAAACACATTCCAACCGGATTTGGCAACGAGCGATTAACTCCAACGCAATCAGC
AACGAGGATTGGTCACAATACTACGGTCTTCTTATAGACCGTTCCCTACTCCATATAAG
GTTAACACAGAGATGGAATGTTCAATGTTACAACACAATATATCTACCACCGTTGCGT
GTAGAGAACTCTGCGATTCTAGTGATTCCGATTTCTTCAGAGTATGACTCACAACAACGAC
CATGGCGTTTTGATGACTTTACTTTGCTGCAAGTAACCTCAACCAACAATAATAGCGTT
GGTGATCAAGTGATCCAGTTGGCAATTATGATGAACAATTAATAACATCTAACCGTCAT
ATGAACCAGACTGGTTATATAAAAGAGCAGAAGATCAGATCGAGTTTGGATAATACTGAC
GAAGATCCAGGATTTTCATGGTAACAATACCAATGACAACATAGATATCGATGATTTTCTC
TCGTTTGATATATATAACGAGGACAACGTGAATCAAATAGAGATAATGAAGACGTGAAT
ACAAATGAAACCCCTGATTTCATCGGGATTGAGGTGGTTGAAGAAGAACTAGATTTAAC
AACC AAATGCTCATCTCGACATATCAAACGACAAAGATTCTATATCACCAGTCGTACCT
TGTCACACGTTGAAAGTTACGTCATCTTATTAGTCACAATGTGGAAGAGAGAACATTG

TTCATTGAAGAGGACAAAGATTCTTGGTTACAAAGAGCTGAGAAGATCACGAAGACAAAA
CTAACACTTTTTAGTTTAAATGGCTCAGCAATACTACAAATGTCTTGCTATTTTTTCTGA
>G2180 Amino Acid Sequence (conserved domain in AA coordinates:7-156)
MAPVSLPPGFRFHPTDEELITYYLKRKINGLEIELEVIAEVDLYKCEPWLPGKSLPSK
DQEWYFFSPRDRKYPNGSRTNRATKGGYWKATGKDRRVSWRDRAIGTKKTLVYYRGRAPH
GIRTGWMHEYRLDETECEPSAYGMQDAYALCRVFKKIVIEAKPRDQHRYSVHAMS NVSG
NCSSSFDTCSLEISSTTHQVQNTFQPRFGNERFNSNAISNEDWSQYYGSSYRPFPTPYK
VNTEIECSMLQHNIIYLPPLRVENSASFSDSDFFTSMTHNNDHGVFDDFTFAASNSNHNSV
GDQVIHVGNVDEQLITSNRHMNQTYIKEQKIRSSLDNTDEDPGFHGNNNTNDNIDIDDFL
SFDIYNEDNVNQIEDNEDVNTNETLDSSGFVVEEETRENNQMLISTYQTTKILYHQVVP
CHTLKVHVNPI SHNVEERTLFI EEDKDSWLQRAEKITKTKLTLFSLMAQQYYKCLAIFF*
>G1817 (1..1308)
ATGAAGGACGCAGAGAAGCGAGAGGTGATTGCATCATCATCATTACAAAGAAAGAGAAAC
AGAGGAAGAAGACTAAGGAAAAAGAAAGAAAGAAACGAGAAGCGAGTACTAATGGTTCCA
TCATCATTACCAAACGACGTGCTAGAGGAGATCTTTTAAAGATTTCCGGTTAAAGCCCTA
ATCCGACTCAAGTCTCTCTCGAAACAATGGAGATCGACGATCGAATCTCGCAGTTTTGAA
GAGAGACACTTGI EDNEDVNTNETLDSSGFVVEEETRENNQMLISTYQTTKILYHQVVP
GGAGAAGAAGATCCCATAGAGGAACCGGGATTTCGTCCAGACACTGACATTTGGTTTTAGG
TTATTCTGCTTGGAATCGGCTTCTCTTCTATCCTTTACTCGTCTCAATTTCCCTCAAGGG
TTCTTCAACTGGATCTACATATCTGAAAGCTGTGATGGCCTTTTCTGCATCCATTCCCCA
AAATCACATTCGGTATATGTAGTGAATCCGGCTACACGGTGGCTCCGCCTACTTCTCTCG
GCAGGGTTTCAGATTTTGATCCACAAGTTTAAACCCCACTGAACGTGAGTGGAAATGTAGTG
ATGAAATCAATCTTTTATCTAGCATTCGTGAAGGCCACCGATTACAAATTAGTGTGGTTG
TACAAATGTGTAGTAAAGTACGTTGATGCGTCGAGTCCAAACGTGGGAGTCACAAAGTGC
GAGATTTTTGACTTTAGGAAAAATGCTTGGAGGTACTTGGCTTGCCTTCAAGTCATCAG
ATATTCTATTACCAAAGCCAGCATCTGCAAACGGGTCGGTTTATTGGTTTACAGAACCA
TATAATGAAAGAATCGAAGTAGTGGCTTTTGATATTCAGACCGAAACATTCCGGTTGCTG
CCTAAGATTAATCCGGCTATTGCTGGTTGAGATCCTCACCATTGACATGTGCACTCTG
GATAATAGTTTGTGTATGTGCGAAAAGGGAGAAAGATACTATGATCCAGATATTTGGAGG
TTGAAACCATCAGAAGACACATGGGAAAAGATTTTTAGCATAGACTTGGTTTCTGTCTCT
TCTTCTCGGACTGAGAAGCGTGATCAATTTGATTGGAGCAAGAAGGATAGGGTTGAGCCA
GCCACACCCGTCGCGGTTTGTAAGAATAAGAAGATCCTTCTCTCACATCGCTATTCGGA
GGTTTGGTAAAGTACGATCCCCTAACAAAATCTATCGATTTTTTTTTCCGGACATCCTACC
GCTTACAGAAAAGTTATTTATTTTCAAAGTTTGATATCTCATCTATAA
>G1817 Amino Acid Sequence (conserved domain in AA coordinates:47-331)
MKDAEKREVIASSSLQKRNRGRRLRKRRRRNEKRVLMVPSSLPNDVLEEIFLRFVVKAL
IRLSLSKQWRSTIESRSFEERHLTIKAKAFVDHPKVMLVGEEDPIRGTGIRPDTDIGFR
LFCLESASLLSFTRLNFPQGFNWIIYISESCDGLFCIHSPKSHSVYVNPATRWLRLLPP
AGFQILIHKNPTEREWNVVMKSI FHAFVKATDYKLWVLYNCDKYIVDASSPNVGVTKC
EIFDFRKNAWRYLACTPSHQIFYYQKPASANGSVYWFTEPYNERIEVVAFDIQETFRLL
PKINPAIAGSDPHHIDMCTLDNSLCMSKREKDTMIQDIWRLKPS EDTWEKIFSIDLVSCP
SSRTEKRDQFDWSKKDRVEPATPVAVCNKNKILLSHRYSRGLVKYDPLTKSIDFFSGHPT
AYRKVIYFQSLISHL*
>G1649 (61..1311)
ATTCACAAAAACCGGAAAAAAGACAAGTAAAGAAAGCTTTGTTCAAGTTTACTTCA
ATGGAAGCAAAACCCCTAGCATCATCATCTGAACCAACATGATTTCTCCATCATCA
AACATTAACCAAAATTAAGAGATGAAGATTATATGGAGCTGGTGTGTGAAAATGGGCAG
ATTCTTGCAAAGATFCGAAGACCAAGAACAACGGTTCTTTTCAAAGCAACGTAGGCAA
TCTCTCCTGGATTTGTATGAGACCGAGTACAGCGAGGGTTTCAAGAAAAACATCAAGATT
CTTGAGACACACAAGTTGTTCCGGTGAGTCAGTCTAAGCCACAACAAGATAAAGAAACC
AATGAACAAATGAACAACAATAAGAAGAAGCTAAAGTCTCCAAAATCGAATTTGAGAGA
AATGTTTTCGAAAAGCAACAATGTGTGAATCATCAACATTAATTGATGTTTCTGCTAAA
GGTCCAAAGAATGTTGAAGTTACTACAGCTCCTCCTGATGAGCAATCTGAGCTGTGGT
AGATCCACGGAATTGATTTTGTCTTCTCATCGAAGTTTTCTCGAGGAACCTCGAGAGAT
CTAAGTTGTTGTTCTTTAAAGAGGAAGTATGGAGATATTGAAGAAGAAGATCAACCTAT
TTAAGTAATAATTGAGATGATGAATCAGATGATGCGAAGACACAAGTTTCATGCGAGAACA
AGAAAGCCGGTGACTAAAAGAAAACGAAGCACAGAAGTCCATAAGTTATATGAAAGAAAA

CGAAGAGATGAATTCAACAAGAAAATGCGTGCTTTGCAGGACCTACTACCAAATTGTTAC
AAGGATGATAAGGCTTCATTGTTGGATGAGGCTATCAAATATATGCGGACCCTTCAACTT
CAAGTTCAGATGATGAGTATGGGAAATGGATTAATAAGACCACCTACGATGTTGCCAATG
GGTCATTACTCTCCCATGGGTCTAGGAATGCATATGGGTGCAGCAGCAACACCAACATCA
ATACCGCAATTCTGCCTATGAATGTTCAAGCAACCGGTTTTCCGGGGATGAACAATGCA
CCACCACAAATGCTAAGCTTTCTTAATCACCCAAGTGGACTAATCCAAACACTCCTATC
TTTTCTCCATTGGAAAATTGCTCTCAGCCATTGCTGGTGCCTTCGTGTGTTTTCTCAGACT
CAGGCTACTTCTTTTACTCAATTTCCCAAAGTCTGCGTCCGCCTCAAACCTAGAAGATGCA
ATGCAATATAGAGGAAGCAACGGTTTTAGTTATTATCGCTCGCCAAACTAATGATTTGTA
GAAAGTTGATGTTTTCTCCAATACTAAGCAAAAAAATGATCGTCTACTCT
GTGTGTTAGTCTATGGGCTTTTGGGCTTGATTTCTTGAACGATTTGAACCTAATTCCA
ACTATTTTCAAAGTGGATGTACAAAGTAAAA

>G1649 Amino Acid Sequence (conserved domain in AA coordinates:225-295)

MEAKPLASSSSEPNMISPSSNIKPKLKDEDYMLVCENGQILAKIRRPKNNGSFQKQRRQ
SLLDLYETEYSEGFKNILKILGDTQVVPVSQSKPQDKETNEQMNNKKKLKSKIEFER
NVSKSNKCVESSTLIDVSAKGPKNVEVTTAPPDEQSAAVGRSTELYFASSSKFSRGTSRD
LSCCSLKRKYGDIEEEESTYLSNNSDDESDAKTQVHARTRKPVTKRKRSTEVHKLIERK
RRDEFNKMRALQDLLPNCYKDDKASLLDEAIKYMRTLQLOVQMMSMGNGLIRPPTMLPM
GHYSPMGLGMHMGAAATPTSIPQFLPMNVQATGFPGMNAPPQMLSFLNHPISGLIPNTPI
FSPLENCSQPFVVPSCVSQTQATSFTQFPKSASASNLEDAMQYRGSNGFSYYRSPN*

>G2131 (69..1010)

GTCTCTCATTTTTCATAATTCATTTCAGGATTGTCTCTCAATCTTTTATTCTTCTCATT
CACCGGTAATGGCAAAAGTCTCTGGGAGGAGCAAGAAAACAATCGTTGACGATGAAATCA
GCGATAAAACAGCGTCTGCGTCTGAGTCTGCGTCCATTGCCCTAACATCCAAACGCAAAC
GTAAGTCGCCCGCCTCGAAACGCTCCTCTTCAACGCAGCTCCCTTACAGAGGCGTCACAA
GGCATAGATGGACTGGGAGATACGAAGCGCATTTGTGGGATAAGAACAGCTGGAACGATA
CACAGACCAAGAAAGGACGTCAAGTTTATCTAGGGGCTTACGACGAAGAAGAAGCAGCAG
CACGTGCCTACGACTTAGCAGCATTGAAGTACTGGGACGAGACACACTCTTGAACCTCC
CTTTGCCGAGTTATGACGAAGACGTCAAAGAAATGGAAGGCCAATCCAAGGAAGAGTATA
TTGGATCATTGAGAAGAAAAGTAGTGGATTTTCTCGCGGTGTATCAAAATACAGAGGCG
TTGCAAGGCATCACCATAATGGGAGATGGGAAGCTAGAATTGGAAGGGTGTGTTGGTAATA
AATATCTATATCTTGAACATACGCCACGCAAGAAGAAGCAGCAATCGCCTACGACATCG
CGGCAATAGAGTACCGTGGACTTAACGCCGTTACCAATTTTCGACGTGAGCCGTTATCTAA
ACCTTAACGCCCGCGCGGATAAAGCCGATTCCGATTCTAAGCCCATTCTGAAGCCCTAGTC
GCGAGCCCGAATCGTCCGATGATAACAAATCTCCGAAATCAGAGGAAGTAATCGAACCAT
CTACATCGCCGGAAGTGATTCCAACCTCGCCGGAGCTTCCCCGACGATATCCAGACGTATT
TTGGGTGTCAAGATTCCGGCAAGTTAGCGACTGAGGAAGACGTAATATTCGATTGTTTCA
ATTCTTATATAAATCTGGCTTCTATAACGAGTTGATTATGGACCTTAATCGTATTTTC
TACAAGTTTTGTTTTGATTATCTACACAATACATCAATATATTTCT

>G2131 Amino Acid Sequence (conserved domain in AA coordinates:50-186, 112-183)

MAKVSGRSKKTIVDDEISDKTASASESASIALTSKRKRKSPPRNAPLQRSSPYRGVTRHR
WTGRYEHLWLDKNSWNTQTKKGRQVYLGAYDEEEAAARAYDLAALKYWGRDILLNFPLP
SYDEDVKEMEGQSKEEYIGSLRRKSSGFSRGSVSKYRGVARHHNGRWEARIGRVFGNKYL
YLGTYATQEEAAIAYDIAAIEYRGLNAVTFNFDVSRYLNPNAADKADSDSKPIRSPSREP
ESSDDNKSPKSEEVIEPSTSPVPIPTRRSFPDDIQTYFGCQDSGKLATEEDVIFDCFNSY
INPGFYNEFDYGP*

>G215 (1..1110)

ATGACTCGTTCGGTTCGTCATTGTAGCAACAATGGGCACAATTCACGCACGTGTCCAACG
CGTGGGTCTGGTTCTCTCCTCCGCCGTGAAGTTATTGGTGTGAGGTTAACGGATGGCTCG
ATTATTAAGAAAGAGTGCAGATATGGGTAATCTCTCGGCATTGCTGTTGCGGCGCGCGG
GCAACGCACCAACCGTTATCTCCGTCGTCTCTCTGGCGACGTCAAATCTTAATGATTGCG
CCGTTATCGGATCATGCCCCGATACTCTAATTTGCATCATAATGAAGGGTATTTATCTGAT
GATCTCTGCTCATGGTTCTGGGTCTAGTCACCGTCGTGGTGAGAGGAAGAGAGGTGTTCTCT
TGGACTGAAGAGGAACATAGACTATTCTTAGTCGGTCTTCAGAAACTCGGGAAAGGAGAT
TGGCGCGGTATTTGAGAAACTATGTAACGTCAAGAACTCTTACACAAGTGGCTAGTCAT
GCTCAAAAGTATTTTATTCGACATACTAGTTCAAGCCGCAGGAAAAGACGGTCTAGCCTC
TTCGACATGGTTACAGATGAGATGGTAACCGATTATCGCCAACACAGGAAGAGCAGACC

TTAAACGGTTCCTCTCCAAGCAAGGAACCTGAAAAGAAAAGCTACCTTCCTTCACTTGAG
CTCTCACTCAATAATACCACAGAAGCTGAAGAGGTCGTAGCCACGGCGCCACGACAGGAA
AAATCTCAAGAAGCTATAGAACCATCAAATGGTGTTCACCAATGCTAGTCCCGGGTGGC
TTCTTTCTCCTTGTTCCTCAGTGACTTACACGATTTGGCTCCCTGCGTCACTTCACGGA
ACAGAACATGCCTTAAACGCTGAGACTTCTTCTCAGCAGCATCAGGTCCTAAACCAAAA
CCTGGATTGCTAAAGAACGTGTGAACATGGACGAGTTGGTCGGTATGTCTCAGCTTAGC
ATAGGAATGGCGACAAGACACGAAACCGAACTTCCCCTTCCCGCTATCTTTGAGACTA
GAGCCCTCAAGGCCATCAGCGTTTCACTCGAATGGCTCGGTTAATGGTGCAGATTTGAGT
AAAGGCAACAGCGCGATTTCAGGCTATCTAA
>G215 Amino Acid Sequence (domain in AA coordinates: TBD)
MTRRCSHCSNNGHNSRTCPTRGSGSSSAVKLFGVRLTDGSIKKASMGNLSALAVAAAA
ATHHRLSPSSPLATSNLNDSPSLDHARYSNLHHNEGYLSDDPAHSGSGSSHRRGERKRGVP
WTEEEHRLFLVLGLQKLGKGDWRGISRNYVTSTPTQVASHAQKYFIRHTSSSRKRSSL
FDMVTDEMVIDSSPTQEEQTLNGSSPSKEPEKSYLPSLELSLNNTEAEVVATAPRQE
KSQEAIEPSNGVSPMLVPGGFFPPCFPVTYTIWLPASLHGTEHALNAETSSQQHQVLKPK
PGFAKERVNMDLGVMSQLSIGMATRHETETSPSPSLRLLEPSRPSAFHSNGSVNGADLS
KGNSAIQAI*
>G1508 (1..420)
ATGCTAGATCACAGTGAAGAGTCTTATTGGTTGATTGAGAAACCATGAAAACAAGAGCT
GAAGATATGATCGAACAGAACAACTAGTGTAAACGACAAGAAGAGACTTGTGCTGAT
TGTGGAACCAAGTAAACTCCTCTTTGGCGTGGTGGTCTGTTGGTCCAAAGTCGTGTGT
AACGCGTGTGGGATCAGAAACAGAAAGAGAAGAGGAGGAACAGAGATAATAAGAAA
TTAAAGAAATCGAGTTCTGGCGGCGAAACCGTAAATTTGGTGAATCGTTAAACAGAGT
TTGATGGATTTGGGGATAAGGAAGAGATCAACGGTGGAGAAGCAACGACAGAGCTTGGT
GAAGAGAACAAGCCGTGTGTTACTCATGGCTCTTTCTTATGGCTCTGTTTACGCTTAG
>G1508 Amino Acid Sequence (domain in AA coordinates: 38-63)
MLDHSEKVLVDSETMKTRAEDMIEQNNTSVNDKKKTCADCGTSKTPLRGGPVGPKSLC
NACGIRNRKRRGGTEDNKKLKKSSSGGGRKFGESLKQSLMDLGIRKRSTVEKQRQKLK
EEEQAAVLLMALSYGSVYA*
>G2110 (36..1622)
GAGAGCTAATAAAAAATTTATCAAGAAGACTAATATGGAGAAGGACGATTTCTTGAGGA
GTGGTCATGGAAGAGAAGAAAGCCATGATGAGATGAGAAAACCTGATTCTCTCACGATG
ATTCTCATCAAGAACACGACCATATTATAAGATCCAAGTTGGACTCAACTAAAGTCGAAA
TGGATGAGGCTAAAGAGGAAAAATCGAAGACTAAAGTCATCATTGAGTAAATCAAGAAAG
ATTTTGACATCCTTCAAACACAATACAACCAATTAATGGCCAAACATAACGAACCAACCA
AGTTCCAATCAAAAAGGGCATCATCAAGACAAAGCGAAGATGAAGACAGAGAAAAAGTTA
ACGAACGTGAAGAAGTGTCTCGTTGAGCCTAGGCGAGACGGTTAAATTCAGAGGTTCCAA
GTGGTTCGAATAAAGAAGAAAAAATAAGATGTTGAAGAAGCGGAAGGTGACAGAAAT
ATGATGATAATGAAAAAAGCAGTATTCAAGGGTTGAGTATGGGGATTGAATACAAGGCTT
TGAGTAATCCTAATGAGAAGTTAGAGATTGATCATAATCAAGAAACCATGTCGTTGGAGA
TTAGTAACAATAAAGATCAGATCAGAAATAGTTTTGGGTTTAAAGAAATGATGGAGATG
ATCATGAAGATGAAGATGAGATTTGCGCTCAAAACCTTGTTAAGAAAACCTAGGGTTTCGG
TGAGATCAAGATGTGAGACACCAACGATGAACGACGGATGTCAATGGAGGAAATATGGCC
AGAAAATAGCTAAAGGCAATCCATGTCCCCGAGCTTACTATCGTTGCACCATTCAGCTT
CTTGTCAGTAAGAAAACAGGTGCAAAGATGTTCAAGAAGATATGTCTATACTTATCTCAA
CGTACGAAGGAACACATAACCATCCACTTCCCATGTCAGCAACTGCCATGGCCTCTGCCA
CTTCCGCTGCCGCTCCATGCTTCTCTCCGCGCCTCTCCTCTCATCCGCCGCGAGCTG
ATCTTCATGGCCTTAACCTTCTCTTTCCGCAACAACATCACTCCAAAACCTAAACTC
ATTTCTCCAATCCCCCTTCTTCTTGGCCATCCGACCGTCACTCTCGACCTCACAACT
CCTCCTCGTCCGACGAACCGTTCTTATCAATGCTCAATAGATTGAGCTCTCCTCCAAGTA
ATGTCTCAGCATCTAATAGTTATCTTCAACCAATCTCAACTTTTCAAACAACCAACA
CATTGATGAATTGGGGTGGTGGTGAATCCAGTGATCAATACCGTGCAGCTTACGGCA
ACATTACACCCATCAGCAATCACCTTACCACAAAATCATTCAAACCCGAACCGCGGGT
CATCTTTCGATCCGTTTGAAGATCATCTTCATCACATTCTCCACAAATAAATCTTGATC
ATATCGGAATCAAGACATCATCAGTCACCAAGTGCATCTTTACCGGCTGAAACAATCA
AGGCAATCAGCAGATCCAAGTTTCCAATCGGCTTTGGCGACAGCTCTATCTTCCATCA
TGGGCGGCGATTAAAGATTGATCACAATGTGACTAGAAATGAAGCTGAGAAGAGCCCTT

AAAGAGAATTGTTATATATATGTTCTTATATACTCAGTACATTGGTAAATGGGTTTAGAC
TTTCACTAGTTTCCTAGTTCATCTATATATTGGTTGTTTAAATCACAAGTTTATTTTGTG
TTGGAGTTTATGGAATAATGTGTACATATGAACTTTAGAACGAATAAATAAACTTGG
AATTCCTTTTTAAAAA

>G2110 Amino Acid Sequence (conserved domain in AA coordinates: 239-298)

MEKDDFLRSGHGREESHDEMRKLDSSHDDSHQEHDI IRSKLDSTKVEMDEAKEENRRLK
SSLKIKKDFDILQTOYNQLMAKHNEPTKFQSKGHHQDKGEDEDREKVNREELVSLSLG
RRLNSEVPSGSNKEKNKDVVEAEEDRNYDDNEKSSIQGLSMGIEYKALSNPNEKLEIDH
NQETMSLEISNNKIRSONSFQKNDGDDHEDEDEILPQNLVKKTRVSVRSRCETPTMND
GCQWRKYGQKIAKGNPCPRAYRCTIAASCPVRKQVQRCSIEDMSILISTYEGTHNHPLPM
SATAMASATSAASMLLSGASSSSAAADLHGLNFSLSGNNITPKPKTHFLQSPSSSGHP
TVTLDLTSSSSSQPFSLMLNRFSSPPSNVSRNSYPSNLFNSNNTNTLMNWGGGNPS
DQYRAAYGNINTHQQSPYHKIIQTRTAGSSFPFRSSSSHSPQINLDHIGIKNIISHQV
PSLPAETIKAITDPSFQALATALSSIMGGDLKIDHNVTREAEKSP*

>G2442 (71..997)

TCGACCAATTTAGACCATTCCAAATTCGTCGTCCTTTTCTCTGTGTAGTCTAATTATATA
TTACAAGTAGATGAATTGGTTACCTGAAGCTGAAGCTGAGGAGCACCTTGAAAGGTATTCT
CTCTGGTGATTCTTTTGTAGGTCTCACCACCTTGATTGCCACTTGAAAGACATCGA
TTCCACCAATGGTGAGGGAGATTGGGTGCGCCAGGTTTCAAGACCTTGAGCCTCCTCCCTT
GGATATGTTCCCTGCTTTGCTTCTGACCTCACCTCTTGTCCCAAGGCGCCGCTCGTGT
GCGGATTCCCAACAACATGATTCCTGCTTTGAAGCAGTCTGTTCTTCTGAAGCCTTGTCT
CGGCATTATAGCACTCCCAACCAATCTTCAGCTCCTCCTGATATCAAAGTTTCATATCT
ATTTCACTCTAATCCAGTGTGAGTTCTCGAGAAGCTTATGGTTCTCTCTCCACCCA
AACTCCGGATCTCAGAGATTGGCTTTCCCTGTGAAGGCATGAGAAGCAAGCGCAGACG
CCCCAACAGCTGAGACTTAGCTACCTTTTCCCTTTGAAGCCAGAAAGTCAACTCCGGG
TGAATCAGTAACCGAGGTTACTATTCTTCTGAGCAACATGCCAAGAAGAAGCGCAAGAT
TCATCTGATCAACACACCGAGTCTTCCACTTTGGAGTCAAGTAAGTCGGATGGGATAGT
CCGGATATGCACTCATTGTGAGACAATCACGACCCACAGTGGAGGCAAGGACCCAGTGG
ACCCAAGACCTCTGCAACGCTTGGCGAGTCCGGTTCAAATCTGGTCGCCTAGTTCCAGA
ATACCGGCCAGCCTCAAGCCCGACCTTCATCCCATCTGTGCATTCAAACCTCACACAGGAA
GATCATTGAGATGAGAAAGAAGGACGACGAGTTTGATACCAGCATGATTGCGAGTGATAT
CCAGAAGGTAAGCAGGGGAGGAAGAAATGGTATAAAAGTA

>G2442 Amino Acid Sequence (domain in aa coordinates: 220-246)

MNWLPEAEAEHLKGLSGDFFDGLTNHLDCPLEDIDSTNGEGDWVARFQDLEPPPLDMF
PALPSDLTSCPKGAARVRIPNNMIPALKQSCSSEALSGINSTPHQSSAPPDIKVSYLFQS
LTPVSVLENSYGLSTQNSGSQRLAFPVKGMRSKRRRPTTVRLSYLFPFEPKSTPGESV
TEGYYSSEQHAKKKRIHLITHTESSTLESSKSDGIVRICTHCETITTPQWRQGPSGPKT
LCNACGVRFKSGRLVPEYRPASSPTFIPSVHSNSHRKIIEMRKODEFDTSMIRSDIQKV
KQGRKKMV*

>G1051 (66..1031)

CCTGTAAATTCAGATTTGCTTTCTTTGGTAATCTTTGGATCAAGATCCATCTATTTTTT
CTTCAATGGCACAACCTCCCTCCTAAAAATCCCAACATGACACAACATTGGCCTGATTCT
CTTCCCAAAGCTCTCTCTTTCTTACCCCAACCGCAACCGCTGTGCGCCACCGCTACAA
CCACCGTACAAAACCCCTCATGGGTGACGAATTCCTCGACTTCTCAGCGTCTCGCCGTG
GCAACCACCGTCTTCATCAGCGACTCTATCGCATTCTCGAAGCTCCAACAGTCAGCA
TCGAAGACCACCAATTGACAGGTTTCGATGACGAACAGTTTCATGTGATGTTACCGGACG
ACGACAACCTTCATAGCAATCTTCCCATATCAACAACAAAAATAACAATGTGGGGCCCA
CGGGATCTTCTCGAACACATCCACGCCGTCCAATAGCTTCAACGACGATAACAAAGAAT
TACCACCGTCCGATCATAACATGAACAATAATATCAACAACAACTATAACGATGAAGTCC
AAAGCCAATGCAAGATGGAGCCAGAAGATGGTACGGCGTGAATAACAATTCGGGTGATA
GCTCCGGCAACCGGATTCTCGATCCCAAAAGGTTAAGAGAATATTAGCAAAATCGGCAAT
CAGCACAGAGATCAAGGTGAGGAACTGCAATACATATCAGAGCTCGAACGTAGCGTCA
CTTCGTTGCAAGCGGAAGTGTGAGTGTATCGCCAAGAGTTGCATTCTTGGATCATCAAC
GTTTGCTTCTTAACGTTGACAAACAGCGCTCTCAAGCAACGAATCGTGCTTTATCTCAAG
ACAAGCTTTTCAAAGACGCACATCAAGAAGCATGAAGAGAGAAATAGAGAGACTTCGAC
AAGTGTATAATCAACAAAGCCTCACGAATGTGGAATGCAAAATCATTTATCGGCGACCG
GAGCCGGTGCTACTCCGGCCGTGACATCAAGTCGTCCTTGAACAGAGCAGCTCCTCA

ATGTCTCATAAATTAACCATCATGCATCATCAACATTTCTCTCTTTAGCTTCTTGG
CAAAAGTTCTTGACTATAAAATCTCTTTCGGGTAAGAAATTCAGGAGATATACATTTTTT
ATTCTAATCACATTGTTTTTAAGTTGTGATGAATTCAGTTTGATGTATCTTATTTATTTT
GTTTATGTCGTCTTTTTTCTTGGGGTTGATGGAAGGAATCATCAATTGTTGTTGTAC
AAAGAACTAGTTGAATTTTTTTTTTTTTTTT

>G1051 Amino Acid Sequence (domain in AA coordinates 189-250)
MAQLPPKIPNMTQHWPDPFSSQKLSFPSTPTATAVATATTTVQNPSWVDFLDFSASRRGN
HRRSISDSIAFLEAPTTSIEDHQFDRFDEQFMSMFTDDNLHSNPSHINNKNNNVGP
SSSNTSTPSNSFNDDNKELPPSDHNMNNINNNYNDEVQSQCMEPEDGTASNNSGDSS
GNRILDPKRVKRILANRQSAQRSRVKRLQYISELERSVTSLQAEVSVLSPRVAFLDHQRL
LLNVDNSALKQRIAAALSQDKLFKDAHQEALKREIERLRQVYNQQLTNVENANHLSATGA
GATPAVDIKSSVETEQLLNVS*

>G1052 (138..1127)

TGATCATCTAAACTTTCAATTTCTCTCTTGATCCTCACTTGAATTTTTTGTGTTTCTC
TCAAATCTTTGATCCTTTCTTTGTTTTTCATTTGACCTCTTACAAAAAATCTGGTGTG
CCATTAAATCTTTATTAATGGCACAACCTTCCTCCGAAAATCCCAACCATGACGACGCCAA
ATTGGCCTGACTTCTCTCCAGAACTCCCTTCCATAGCCGCAACGGCGGCAGCCGCAG
CAACCGTGGACCTCAACAACAAAACCTTCATGGATGGATGAGTTTCTCGACTTCTCAG
CGACTCGCGTGGGACTCACCGTCGTTCTATAAGCGACTCCATTGCTTTCTTGAACCAC
CTTCTCCGGCGTCGGAAACCACTTCGATAGGTTTGACGACGAGCAATTCATGTCCA
TGTTCAACGACGACGTACACAACAATAACCACAATCATCATCATCACAGCATCAACG
GCAATGTGGGTCCACGCGTTTCATCTCCAACACCTCCACGCGTCCGATCATAATAGCC
TTAGCGACGACGACAACAACAAGAAGCACCACCGTCCGATCATGATCATCATGAGACA
ATAATGTAGCCAATCAAAACAACGCCCGCGGTAACAATTACAACGAATCAGACGAGGTCC
AAAGCCAGTGCAAGACGAGCCACAAGATGGTCCGTCGGCGAATCAAACTCCGGTGGAA
GCTCCGTAATCGTATTCACGACCCCTAAAAGGGTAAAAGAATTTAGCAAATAGGCAAT
CAGCACAGAGATCAAGGGTGAGGAAATTGCAATACATATCAGAGCTTGAAAGGAGCGTTA
CTTCATTGCAGACTGAAGTGTGAGTTATCGCCAAGAGTTGCGTTTTTGGATCATCAGC
GATTGCTTCTCAACGTCGACATAGTGCTATCAAGCAACGAATCGCAGCTTTAGCACAAG
ATAAGATTTTCAAAGACGCTCATCAAGAAGCATTGAAGAGAGAATAGAGAGACTTCGAC
AAGTATATCATCAACAAAGCCTCAAGAAGATGGAGAATAATGTCTCCGATCAATCTCCGG
CCGATATCAACCGTCCGTTGAGAAGGAACAGCTCCTCAATGTCTAAAGCTGTTCTGTTCA
CTAAGATCTTTCTTTTCATGGCGAAAAGATTCTTGACTATAAAACCTCTTTGTGTCAAGA
AATTAATTTATCAAGAAGATGGCCTTTTTTATTTGATCTAATCACATTTTTTTAAGTTG
TGATGAATTTGCTTTTGATGTATCTGTTTTTTTTTTTTTTTTT

>G1052 Amino Acid Sequence (domain in AA coordinates 201-261)
MAQLPPKIPTMTTPNWPDPFSSQKLPISIAATAAAAATAGPQQNPSWMDLDFSATRRGT
HRRSISDSIAFLEPPSSGVGNHHRFDRFDEQFMSMFDNDVHNNNNHHHHHSINGNVGPT
RSSSTSTPSDHNLSDDDNKEAPPSDHDHMDNNVANQNNAAGNNYNESDEVQSQCKT
EPQDGPANQNSGSSGNRIHDPKRVKRILANRQSAQRSRVKRLQYISELERSVTSLQTE
VSVLSPRVAFLDHQRLLLNVDNSAIKQRIAAALAQDKIFKDAHQEALKREIERLRQVYHQ
SLKKMENNVSQSPADIKPSVEKEQLLNVS*

>G1079 (1..1995)

ATGGGTTGTGCTGCTTCAAGAATTGATAATGAAGAAAAGGTTTATAGTGTGTAGGCAGAGA
AAGAGGCTAATGAAAAAGTTATTAGGGTTCAGGGGAGAATTGTCAGATGCACAGTTGGCT
TATCTTAGAGCTTTGAGGAACACTGGTGTACTCTTAGGCAATTCAGTGTCTGAGACC
TTGGAGCTTGAAAACTAGTTATGGTTTAAAGTTGCTTTGCCTCCTTCGCCTCCTCCT
ACATTGCCTCCTTCACCTCCACCACCTCCTCCATTTAGCCCGGATTTGAGAAATCCTGAG
ACTAGTCATGACTTGGCTGATGAGGAGGAAGAGGGTGAAAATGATGGTGGTAATGATGGA
AGTGGTGCAGCTCCTECGCCTCCATTGCCGAATCTTGGAACATTTGGAACCCCTTTGAG
TCACTTGAGCTGCATAGTCATCCAAATGGTGACAATGTAGTTACACAAGTTGAACTGAAG
AAGAAACAACAATTCAGCAAGCTGAAGAGGAAGATTGGCGGAGACGAAGTCTCAATTT
GAGGAAGAAGATGAGCAACAAGAAGCAGGAGGTACTTGCCTTGATTTGAGTGTTCATCAA
ATAGAGGCTGTTAGTGGCTGTAAACATGAAGAAGCCACGTCGTCTGAAGTTAAGCTGGGA
GAAGTTATGGACGGTAACATCTATGACAAGCTGCTCCGGTAAAGATCTTGAGAAAACCT
CATGTGACTGATTGTAGAATCAGGAGGACCTTAGAAGGAATCATCAGAGAGTTGGATGAT
TATTTTCTTAAAGCATCGGGTTGCCGAGAAGGAGATAGCTGTGATAGTAGACATCAACAGT

AGGGATACTGTTGATCCTTTTCAGGTACCAGGAAACAAGAAGGAAGAGAAGCAGCTCGGCA
AAGGTATTCACTGTCATTGTTCATGGAGTTGGTCTTCAAAGTCTCTTCAGTTGGGCAAAGAT
GCTACAACAAGCGGGACTGTTGAACCTGTAGGCCTGGAGCTCACTGCAGCACACTTGAG
AAGCTATACACAGCTGAGAAGAACTTTACCAGCTAGTCAGAAACAAAGAGATTGCCAAA
GTGGAGCATGAGAGGAAGTCTGCATTACTGCAAAAGCAAGATGGGGAAACCTATGATTTG
AGCAAAATGGAGAAAGCACGCTTGTCTTTGGAGAGTTTGGAAACCGAGATACAGCGTCTA
GAAGATCCATAACTACAACAGCTCATGTTTGCTTAACCTTGATCAATGATGAGCTGTAT
CCGCAGCTAGTTGCTTTAACTTCAGGGCTAGCACAGATGTGGAAAACAATGCTCAAGTGT
CATCAAGTTCAAATTCATATATCCCAGCAACTGAACCATCTTCCGGATTACCCGAGTATA
GATCTCAGTTCGGAATACAAACGCCAGGCGGTTAATGAACTAGAGACCGAGGTTACTTGC
TGGTACAATAGCTTTTGCAAGTTAGTAAATTCACGAGGAGAATACGTGAAAACACTCTGT
ACGTGGATCCAACCTTACTGATCGCCTCTTAACGAAGACAACCAAAGAAGTAGCTTGCCT
GTTGCTGCTCGTAAGCTCTGCAAGAGTGGCAGCTTGAATACAACCTGCGTAGGAAATGC
AATAAACTTGAGAGGAGGCTTGAGAAAGAGCTAATTTCACTGGCTGAGATTGAAAGAAGG
CTCGAGGGGATTTTAGCAATGGAAGAGGAGGAAGTAAAGCTCAACGAGTTTGGGCTCTAAG
CATCCGTGTCAATCAAACAAGCCAAGATCGAAGCCTTGAGAAAACGAGTGGATATTGAG
AAAATAAGTACTTAACTCGGTCGAGGTTAGTAAGAGAATGACACTAGACAACCTCAA
TCAAGCCTTCCCAATGTCTTTCAGATGTTGACTGCTCTAGCTAATGTCTTTGCCAATGGG
TTTGAATCCGTTAATGGCCAAACCGGTACAGATGTTTCCGACACATCCCAACATTCCGAT
GAATCTCAACCCTAA

>G1079 Amino Acid Sequence (conserved domain in AA coordinates:1-50)

MGCAASRIDNEEKVLVCRQRKRLMKLLGFRGEFADAQLAYLRALRNTGVTLRQFTESET
LELENTSYGLSLPLPPSPPTLPSPPPPPFSPDLRNPETSHDLADEEEEEEGENDGGNDG
SGAAPPPLPNSWNIWNPFFESLELHSHPNGDNVVTQVELKKKQQIQAAEEDWAETKSQF
EEDEQQEAGGTCLDLSVHQIEAVSGCNMCKPRRLKFKLGEVMDGNSSMTSCSGKDELEKT
HVTDCRIRRTLEGI IRELDDYFLKASGCEKEIAVIDINSRDTVDPFRYQETRRKRSSSA
KVFSALSWSWSSKSLQLGKDATTSGTVEPCRPGAHCSTLEKLYTAEKKLYQLVRNKEIAK
VEHERKSALLQKQDGETYDLSKMEKARLSLESLETEIQRLSDSITTRSCLLNLINELEY
PQLVALTSGLAQMWKMLKCHQVQIHISQQLNHLDPYPSIDLSSEYKQAVNELETEVTC
WYNSFKLVNSQREYVKTLCTWIQLTDRLSNEDNQRSSLPVAARKLCKEWQLEYNLRRKC
NKLERRLEKELISLAEIERRLEGILAMEEEEEVSSTSLGSKHPLSIKQAKIEALRKRV DIE
KTKYLSVEVSKRMTLDNLKSSLPNVFQMLTALANVFANGFESVNGQTGTDVSDTSQHS D
ESQP*

>G1335 (56..667)

TTTTTTTTTAAAAGATTTAGAGAGAAAAGTGAGTTATTAAGAGATTCCAATCAAATGAG
CGGAGACAACGGCGGTGGTGAGAGGCGCAAAGGCTCCGTCAAGTGGTTTGATACCCAGAA
GGGTTTCGGCTTCATCACTCCTGACGACGGTGGCGACGATCTCTTCGTTCAACAGTCCTC
CATCAGATCTGAGGGTTTCCGTAGCCTCGCTGCCGAAGAAGCCGTAGAGTTCGAGGTTGA
GATCGACAACAACAACCGTCCCAAGGCCATCGATGTTTCTGGACCCGACGCGCTCCCGT
CCAAGGAAACAGCGGTGGTGGTTTCATCTGGCGGACGCGCGGTTTTCGGTGGAGGAAGAGG
AGGTGGACGCGGATCTGGAGGTGGATACGGCGGTGGCGGTGGTGGATACGGAGGAAGAGG
AGGTGGTGGTTCGAGGAGGACGCACTGCTACAAGTGTGGTGGAGCCCGTCACATGGCGAG
AGACTGTTCTGAAGGCGGTGGAGGTTACGGAGGAGGCGGCGGTGGCTACGGAGGTGGAGG
CGGATACGGCGGAGGAGGTGGTGGTTACGGAGGTGGTGGCGGTGGAGGTGGTGGCGGCGG
GGGAAGCTGCTACAGCTGTGGCGAGTCCGGACATTTCCGCAGGGATTGCACCAGCGGTGG
ACGTTAAACCAACGCCGGTTACGCGGTGGAGAAGAGTGAGTTGGTTATCTCAAGTGA
TCGGTTCTTTCTCCCGCCGCTTCTATCTCTCTATTATCCACTTTTGTCTATTATGATG
GATCTCTATCTTTGTTAGTTGGTTTTTTCTTGATGGTTTCGGATTAGGACTCTTCTTTTG
GTTTTGCTACTTATGGTTGGTTTTATTTATGGTACTTGTGATATGGGTGAAATGCTCTAC
TTGTTGCTCTGTTTCAAGTGTTCATAATATGCCAACAATATTCTGGGTTTTGTTTCAA
AAAAA

>G1335 Amino Acid Sequence (domain in AA coordinates: 24-43, 131-144, 185-203)

MSGDNGGERRKGSVKWFDTKGFGFITPDDGGDDLFLVHQSSIRSEGRSLAAEEAVEFE
VEIDNNRNPKAIDVSGPDGAPVQNSGGSSGGRGGFGGGRGGGRSGGGYGGGGGGYGG
RGGGGRRGSDCYKCGEPGHMARDCEGGGGYGGGGGGYGGGGYGGGGGGYGGGGRRGGG
GGGSCYSCGESGHFARDCTSGGR*

>G157 (31..621)

GGGCATAACCCCTTATCGGAGATTGAAGCCATGGGAAGAAGAAAAATCGAGATCAAGCGA
ATCGAGAACAAAAGCAGTCGACAAGTCACTTTCTCCAAACGACGCAATGGTCTCATCGAC
AAAGCTCGACAACCTTCGATTCTCTGTGAATCCTCCGTCGCTGTTGTCGTATCTGCC
TCCGGAAAACTCTATGACTCTTCCTCCGGTGACGACATTTCCAAGATCATTGATCGTTAT
GAAATACAACATGCTGATGAACCTTAGAGCCTTAGATCTTGAAGAAAAAATTCAGAATTAT
CTTCCACACAAGGAGTTACTAGAAACAGTCCAAAGCAAGCTTGAAGAACCAATGTCGAT
AATGTAAGTGTAGATTCTCTAATTTCTCTGGAGGAACAACCTTGAGACTGCTCTGTCCGTA
AGTAGAGCTAGGAAGGCAGAACTGATGATGGAGTATATCGAGTCCCTTAAAGAAAAGGAG
AAATTGCTGAGAGAAGAGAACCAGGTTCTGGCTAGCCAGATGGGAAAGAATACGTTGCTG
GCAACAGATGATGAGAGAGGAATGTTCCGGGAAGTAGCTCCGGCAACAAAATACCGGAG
ACTCTCCCGCTGCTCAATTAGCCACCATCATCAACGGCTGAGTTTTCACCTTAAACTCAA
AGCCTGATTCAATAAAGAGAATAAATTTGTATATTATAAAAAGCTGTGTAATCTCAAA
CCTTTTATCTTCTCTAGTGTGGAATTTAAGGTCAAAAAGAAAACGAGAAAGTATGGATC
AGTGTGTACCTCTCTCGGAGACAAGATCAGAGTTTGTGTGTTGTGTCTGAATGTACGG
ATTGGATTTTTAAAGTTGTGCTTTCTTTCTTCAAAAAA

>G157 Amino Acid Sequence (domain in AA coordinates: 2-57)

MGRRKIEIKRIENKSSRQVTFSKRRNGLIDKARQLSILCESSVAVVVVSASGLYDSSSG
DDISKIIDRYEIQHADELRALDLEEKIQNYLPHKELLETVQSKLEEPNVDNVSVDLSLISL
EEQLETALSVSRARKAELMMEYIESLKEKEKLLREENQVLASQMGKNTLLATDDERGMFP
GSSSGNKIPETLPLLN*

>G1895 (1..954)

ATGAATAACCAATCTGTTACTGACAATACAAGTCTTAAGCTGTCATCTAATCTTAACAAC
GAGTCAAAAGAAACATCTGAGAACAGTGTGACCAACACAGCGAGATCACAACAATTACA
TCGGAAGAAGAGAAAACAACCTGAAGTGAAGAAACAGACAAGATTCTTCCATGTCCGAGA
TGCAACAGCGCAGACACCAATTTCTGTTACTACAACAACCTACAACGTTAACCAGCCACGT
CACTTCTGTAGAAAATGCCAGAGGTATTGGACCGCTGGTGGATCCATGAGGATCGTCCCG
GTTGGCTCAGGCCGTCGCAAGAACAAGGGATGGGTTTCTTCAGACCAGTACCTGCACATC
ACTTCCGAGGATACTGACAATTACAATAGCTCCTCAACAAAGATTCTAAGCTTCGAGTCT
TCGGACTCTTTGGTAACTGAGAGGCCCTAAGCATCAATCAAACGAAGTGAAGATAAACGCT
GAACCTGTTTCACAAGAACCCAACTTCCAAGGGTTACTTCTCTCCCAAGCATCCCTT
GTTTCGCCTCCTTGGCCTTACCAATACCCTCCAAACCCTAGTTTCTACCACATGCCCGTC
TACTGGGGCTGCGCGATACCGGTTTGGTCTACCTCGACACTTCTACATGTCTTGGGAAA
AGGACAAGAGACGAACTTCTCATGAAACTGTTAAAGAGAGTAAAAATGCTTTTGAGAGA
ACAAGCTTGCTTTTGAATCTCAGAGCATCAAAAATGAAACAAGTATGGCTACAAATAAC
CATGTGTGGTATCCAGTACCGATGACCCGCGAGAAGACACAAGAATTACGCTTTTTTCAGT
AATGGAGCTGAAACAAAGAGCAGCAACAACAGATTTCGTCCCTGAAACGTATCTTAACCTG
CAAGCAAACCTGCAGCCATGGCAAGATCTATGAACCTCAGAGAGAGCATATAA

>G1895 Amino Acid Sequence (domain in AA coordinates: 55-110)

MNNQSVTDNTSLKLSSNLNNEKETSSENSDDQHSEITITSEEEKTTTELKPKDILPCPR
CNSADTKFCYNNYNVNQPRHFCRKCQRYWTAGGSMRIVPVGSGRRKNKGWVSSDQYLHI
TSEDITDYNSSSTKILSFESSDSLVTTERPKHQSEVKINAEPVSQEPNNFQGLLPQASP
VSPFPWPYQYPNPSFYHMPVYWGCAIPVWSTLDTSTCLGKRTRDETSHETVKESKNAFER
TSLLESQSIKNETSMATNNHVWYVPVMTREKTQEFSSFGAETKSSNNRFPVETYLNL
QANPAAMARSMNFRESI*

>G1900 (1..897)

ATGCTGGAAACTAAAGATCCTGCGATAAAGCTCTTTGGTATGAAAATTCCTTTCCCGACG
GTTTTAGAGGTTGCTGATGAAGAAGAAAAGAACCAAAACAAGACATTAAGTATCAA
TCGGAGAAAAGACAAACCTTAAAGAAACCAACCAAGATTCTTCCATGTCCAAGATGCAAC
AGCATGGAGACTAAGTTCTGTTACTACAACAACCTACAACGTAAACCAACCTCGCCATTTT
TGTAAGCTTGTGATGAGATATTGGACCTCAGGTGGGACCATGAGAAGTGTTCGAATCGGA
GCAGGACGGCGCAAGAACAAGAACTACCAACTTCACATTACCACCATGTGACTATC
TCCGAAACAAATGGTCCGCTTCTAGTTTCAGCCTCGGAGATGATCAAAAGGTCTCGAGT
AATAGGTTTGGTAATCAAAAGCTAGTTGCTAGGATAGAGAACAATGACGAGCGCTTAAT
AACAACACTTCGAACGGTTTGAATTGTTTCCGGGAGTTTCGTGGCCGTACACGTGGAAT
CCTGCGTTTACCCGTTTACCCTTATTGGAGCATGCCAGTGTGCTTCTCCGGTAAAGT
TCAAGTCTACTTCTACTCTTGGTAAGCATTCGAGAGACGAAGACGAGACGGTGAAGCAA
AAACAGAGGAATGGATCTGTATTGGTTCCAAAGACTTTGAGAATTGATGATCCTAATGAA

>G2007 Amino Acid Sequence (domain in AA coordinates: TBD)
MGRQPCCDKLMVKKGPWTAEEDEKILNFI LTNGHCCWRALPKLAGLRRCGKSCRLRWNTY
LRPDLKRGLLSDAEEQLVIDLHALGNRWSKIAARLPGRTDNETIKNHWNTHIKKKLLKME
IDPSTHQLPNKVFTDINLVDKSETSSKADNVNDNKIVEIDGTTTNTIDDSIITHQNSSND
DYELLGDIHNYGDLFINLWNTDEPLVDDASNSHNHVGIGTTAAVAASDKNNTAAEEDF
PERSFEKQNGESWMFLDYCQEFVGVEDFGFECYHGFQSSMKTGHKD*

279

CGGTCTCGTGTGGCTCAAACACTCCGTCGAGTAGTGATGATGTTGAGGCGGATGCATCA
GAAAGGCAAGAGGATGGCACCAATGGTGAGGTGAAAGAAACGAATGAAGACACTAATAAA
CCTCAAACCTTCAGAGTCCAATGCACGCCGAGTAGAATCAGCTCCAATATAACCGATCCA
TGGAAGTCTGTGTCTGACGAGGGTCAATTGCCCTTCCAAGCTCTCTCTCCAGAGAGGTA
TTGCCGCAAAGTTTTACATATCGAGAAGAACAAGAGAGGAAGAACAACAACAAGAA
CAAAGATATCCAATGGCACTTGATCTTAACCTTACAGCTCAGTTAACACCAGTTGATGAT
CAAGAGGAGAAGAGAAACACAGGATTTCTTGAATCGGATTAGATGCTTCAAAGCTAATG
AGTAGAGGAAGAACAGGTTTTAAACCATACAAAGATGTTCCATGGAAGCCAAAGAAAGT
AGAATCCTCAACAACAATCCTATCATTCATGTGGAACAGAAAGATCCCAAACGGATGCGG
TTGGAAACTCAAGCTTCCACATGAGACTCTATTTTCATCTGATCTGTTGTTTGTACTCTG
TTTTTAAGTTTTCAAGACCACTGCTACATTTTCTTTTTCTTTTGAGGCCTTTGTATTGT
TTCCTTGTCATAGTCTTCTGTAAACATTTGACTCTGTATTATTCAACAATCATAACT
GTTTAATCTTTTTTTTTTCCA

>G214 Amino Acid Sequence (domain in AA coordinates: 22-71)
METNSSGDELVIKTRKPYTITKQRRWTEEEHNRFI EALRLYGRAWQKIEEHVATKTAVQ
IRSHAQKFPFSKVEKEAEAKGVAMGQALDIAIPPPRKRKPNPNPYPRKTGSGTILMSKTGV
NDGKESLGSEKVSHPMANEDRQSKPEEKLQEDNCSDCFTHQYLSAASSMNKSCIETS
NASTFREFLPSREEGSQNNRVRKESNSDLNAKSLENGNEQGPQTYPMHIPVLVPLGSSIT
SSLSHPPSEPDSPHTVAGDYQSFNHHMSTLLQTPALYTAATFASSFWPPDSSGGSPVP
GNSPPNLAAMAAATVAASAWWAANGLLPLCAPLSSGGFTSHPPSTFGPSCDVEYTKAST
LQHGSVQSREQEHSEASKARSSLDSEDEVENKSKPVCHQPSATPESDAKSGDAGDRKQV
DRSSCGSNTPSSDDVEADASERQEDGTNGEVKETNEDTNKPQTSSESNNRRSRISSNITD
PWKSVSDEGRIFALFQALFSREVLPSQSFYREEHREEBQQQEQRYPMALDLNFTAQLTPVD
DQEEKRNTGFLGIGLDASKLMSRGRGTGFKPKYKRCMEAKESRILNNNPIIHVEQKDPKRM
RLETQAST*

>G2155 (63..740)
CTCATATATACCAACCAAACCTCTCTCTGCATCTTTATTAAACACAAAATCCAAAAGATT
AAATGTTGTGGAAGCTCCCTACACAGCGACACTTGCACCTCTCTCCCTCTCCCTCCA
TGGAACCGTCGGGCGTCCACGTGGCAGACCTCGAGGTTCCAAAAACAAACCTAAAGCTC
CAATCTTTGTCAACATTGACCCTCCTATGAGTCCTTACATCCTCGAAGTGCCATCCGGAA
ACGATGTCTGTTGAAGCCCTAAACCGTTTCTGCCGCGGTAAAGCCATCGGCTTTTGGCTCC
TCAGTGGCTCAGGCTCCGTTGCTGATGTCACCTTGGCTCAGCCTTCTCCGGCAGCTCCTG
GCTCAACCATTACTTTCCACGGAAGTTCGATCTTCTCTCTGCTCCGCCACTTTCCTCC
CTCCTCTACCTCCTACCTCCTTGTCCCTCCCGTCTCCAATTTCTTACCGTCTCTCTCG
CCGGACCTCAGGGGAAAGTCATCGGTGGATTGCTCGCTGGTCTCTCGTTGCCGCGGAA
CTGTTTACTTCGTCGCCACTAGTTTCAAGAACCCTTCTATCACCGGTTACCTGCTACGG
AGGAAGAGCAAAGAACTCGGCGGAAGGGGAAGAGGAGGGACAATCGCCGCCGGTCTCTG
GAGGTGGTGGAGAGTACGTGCGGTGGCTCTGATGTCATTGGGATCCCAACGCCA
AAGCTCCATCGCCGTACTGACCACAAATCCATCTCGTTCAAACCTAGGGTTTCTTCTTCTT
TAGATCATCAAGAATCAACAAAAGATTGCATTTTATGATTCTTTGTAATATCATAATTG
ACTCACTCTTTAATCTCTCTATCACTTCTTCTTTAGCTTTTTCTGCAGTGTCAAACCTCA
CATATTTGTAGTTTGTATTGACTATCCCAAGTTTGTATTTTATCATACAAATTTTGC
CTGTCTCTAATGGTTGTTTTTCTGTTGTATAATCTTATGCATTGTTTATTGGAGCTCCA
GAGATTGAATGTATAATATAATGGTTAAT

>G2155 Amino Acid Sequence (domain in AA coordinates:18-38)
MLSKLPQQRHLHLSPPSSPMETVGRPRGRPRGSKNKPAPIFVTIDPPMSPYILEVPSGN
DVVEALNRFRCRGAIGFCVLSGSGSVADVTLRQPSAAPGSTITFHGKFDLLSVSATFLP
PLPPTSLSPVSNFFTVSLAGPQGVIGGFVAGPLVAAGTVYFVATSFKNPSYHRLPATE
EEQRNSAEGEESQSPVSGGGGSMYVGGSDVIWDPNAKAPSPY*

>G234 (106..1035)
CACAACATCATACCCACCAACATATATAATCTTGATCATAGAGAGATAAACAGAGGCCGC
TATCAAGAACAAAGACTAAGAACAAAGACTTCACTAGGAGTACAAGTATGGGAAGAGCACCG
TGTTGTGACAAAGCAAACGTGAAGAAAGGGCCTTGGTCTCCTGAGGAAGATGCAAAACTC
AAATCTTACATTGAAAATAGTGGCACCGGAGGCAATTGGATCGCTTTGCCCTCAAAGATT
GGTTTTAAGAGATGTGGAAGAGTTGCAGGCTGAGGTGGCTTAACCTATCTTAGACCAAAC
ATCAAACATGGTGGCTTCTCTGAGGAAGAAGAAAACATATTGTAGCCTTTACCTTACA
ATTGGTAGCAGGTGGTCTATAATCGCTGCTCAATTGCCGGGACGAACAGACAACGATATA

AAAACTATTGGAACACGAGGCTCAAGAAGAACTCATTAACAAACAACGCAAGGAGCTT
CAAGAAGCTTGTATGGAGCAGCAAGAGATGATGGTGATGATGAAGAGACAACACCAACAA
CAACAAATCCAACTTCTTTTATGATGAGACAAGACCAAAACAATGTTTACATGGCCACTA
CATCATCATAATGTTCAAGTTCAGCTCTTTTCAGAATCAAACCAACTCGTTTTGCGACC
AAGAAGATGTTAAGCCAGTGCCTCATCAAGAACATGGTCAAGATCGAAGATCAAGAACTGG
AGAAAACAAACCTCATCATCATCAAGATTCAATGACAACGCTTTTGATCATCTCTCTTTC
TCTCAACTCTTGTAGATCCTAATCATAACCACCTTAGGATCAGGAGAGGGTTTCTCCATG
AACTCTATCTTGAGCGCCAACACAAACTCTCCATTGCTTAACACAAGTAATGATAATCAG
TGGTTCGGGAATTTCCAGGCCGAAACCGTAAACTTGTCTCAGGAGCCTCCACAAGTACT
TCGGCAGATCAAAGCACTATAAGTTGGGAAGACATAAGCTCTCTGTTTATTCTGATTCA
AAGCAATTTTTTAAATTATAATAATATATTATTCTTAAGATGAAACGTACATCATTATTA
TTAATTGGGGTACGTAACGTATATATGGAATAACGATCTAGTTTGTAAATTTAAAA
>G234 Amino Acid Sequence (domain in AA coordinates: 14-115)
MGRAPCCDKANVKGPWSPEEDAKLKSIIENSGTGGNWIAPQKIGLKRCGKSCRLRWLN
YLRPNIKHGGFSEEEENIICSLYLTIGSRWSIIAQLPGRDNDIKYWNTRLKKLINK
QRKELQEACMEQEMVMKQHQQQQIQTSFMMRQDQTMFTWPLHHHNVQVPALFRIKP
TRFATKKMLSQCSSRTWSRSKIKNWRKQTSSSSRFNDNAFDHLSFSQLLDPNHNHLGSG
EGFSMNSILSANTNSPLNLSNDNQWFGNFQAETVNLFSGASTSTSADQSTISWEDISSL
VYSDSKQFF*

>G361 (54..647)

TCTGTCTCTCTCTCTCTCTTTGTAAATATACATATATAGATAAGCTCACATATATGGCGA
CTGAAACATCTTCTTTGAAGCTCTTCGGTATAAACCTACTTGAAACGACGTCGGTTCAAA
ACCAGTCATCGGAACCAAGACCCGGATCCGGATCAGGATCCGAGTCACGTAAGTACGAGT
GTCAATACTGTTGTAGAGAGTTTGCTAACTCTCAAGCTCTTGGTGGTCACCAAAACGCTC
ACAAGAAAGAGCGTCAGCTTCTTAAACGTGCACAGATGTTAGCTACTCGTGGTTTGCCAC
GTCATCATAATTTTACCCTCATACCAATCCGCTTCTCTCCGCCCTTCGCGCCGCTGCCTC
ACCTCCTCTCTCAGCCGCATCCTCCGCCGCATATGATGCTCTCTCTCTCTCTCTCGAGTT
CTAAGTGGCTTTACGGTGAACACATGTCGTCACAAAACGCCGTTGGGTACTTTCATGGTG
GAAGGGGACTTTACGGAGGTGGCATGGAGTCTATGGCCGGAGAAGTAAAGACTCATGGTG
GTTCTTTGCGGAGATGAGGAGTTTCGCCGAGATAGTGATCGGAGTAGCGGAATTAAGT
TAGAGAATGGTATTGGGCTGGACCTCCATTTAAGCCTTGGGCCATGAATGATTATAATTT
TGGCCCAGTAAAGATCTGTAAAATACTACTAGGATTTTCAATTTTATAGAGTATGTTTTTT
TCCTTAATTTCCGTTGAAATTGGTGAATATTTTATCTCTTACTTACCAATCTCATATT
TCTATGTATGCGTTTGGCTTTCATTTTTTTTTTATATAATTCTTCTGTAAAAAATGCA
ATGTGAGTTTTCTCCCTATCATTCTGTCAAGCTTTGGTTCAATTATTTAGTAATCGAAT
AATATAGGAATAGTGTGAAAG

>G361 Amino Acid Sequence (domain in AA coordinates: 43-63)
MATETSSSLKLFGINLLETTSVQNQSSEPRPGSGSGSESRYECQYCCREFANSQALGGHQ
NAHKKERQLLKRAQMLATRGLPRHNFHPHTNPLLSAFAPLPHLLSQPHPPHMLSPSS
SSSKWLYGEHMSSQNAVGYFHGGRGLYGGMESMAGEVKTHGGSLEPMRRFAGDSDRSSG
IKLENGIGLDLHLSLGP*

>G562 (137..1285)

ATTTGAATTTCTGGGTTTCTCTCTGTTTAAAGCTTCTTCTTCTTCTCATCTTCTGCTTACGTT
TCTTCTTCAAGGAGCTTTCGGATTCTTGTAGAAAGAGTCATTGTTCTTCTTGGAGTGGGAAA
CCTTGAAACCATTCCTATGGGAAATAGCAGCGAGGAACCAAAGCCTCCTACCAATCAGA
TAAACCATCTTCACCCCGGTGGATCAAACAAATGTTTATGCTTACCTGATTGGGCAGC
TATGCAAGCATATTATGGTCCAAGAGTAGCAATGCCTCCTTATTACAATTACGCTATGGC
TGCATCTGGTCTCTCTCTCTCTCTTACATGTGGAATCCTCAGCATATGATGTACCATC
TGGAGCACCTATGCTGCTGTTTATCCTCATGGAGGAGGAGTTTACGCTCATCCCGGTAT
TCCCATGGGATCACTGCCTCAAGGTCAAAGGATCCACCTTTAACAACCTCCGGGGACGCT
TTTGAGCATCGACACTCTACTAAATCTACAGGGAACACAGACAATGGATTGATGAAGAA
GCTGAAAGAGTTTGTATGGGCTTGTATGTCTCTAGGAAATGGGAATCCTGAAAATGGTGC
AGATGAACATAAACGATCAGGAACAGCTCAGAAACTGATGGTTCTACTGATGGAAGTGA
TGGGAATACAACCTGGGCGAGATGAACCGAAACTTAAAGAAGTCGAGAGGGAACCCAAC
AAAAGATGGGAACAATGGTTTCAAGCTAGCTCATTTTCTGTTTCTCCGTCAGGTGG
TGATACCGGCGTAAACTCATTCAGGATCTGGAGCTATACCTCTCTCTGTTGTAAGTGC
AAATTCACACCCCTTCATGTCACAATCTTTAGCCATGGTTCTCTCTGAAACTTGCTTCA

GAACGAGAGAGAACTGAAACGGGAGCGAAGGAAACAGTCTAATAGAGAATCTGCTAGAAG
GTCAAGATTAAGGAAACAGGCCGAGACAGAAGAAGCTTGCTAGGAAAGTGAAGCCTTGAC
AGCCGAAAACATGGCATTAAAGATCTGAACTAAACCACTTAATGAGAAATCTGATAAACT
AAGAGGAGCAAATGCAACCTTGTGGACAACTGAAATGCTCGGAACCCGAAAAGAGAGT
CCCCGAAATATGTTGTCTAGAGTTAAGAACTCAGGAGCTGGAGATAAGAACAAAGAACCA
AGGAGACAATGATTCTAATCTACAAGCAAATTCATCAACTGCTCGATACGAAGCCTCG
AGCTAAAGCAGTAGCTGCAGGCTGAATCGATGGTAATTCATGTCGATTCTACTTAATTT
GTCGACATAAACAAAGAAAATAAGTGCTACTAATTTTCAGAAAACTTGATAGATAGATAG
TATAGTAGAGAGAGAGAGAGAGAGAGAGGCTGTGATGATTATTGATCTATAAATTTTCGGA
GAGAGAGAGGGAGAAAGAGAACTTTTCTCCAGATGAAATTTGGTGTTATGGTTTGT
ACTGTTAATATAGAGAGGCTTTTCTTTTATAAAATGGCTTCTTTGTTGCA

>G562 Amino Acid Sequence (domain in AA coordinates: 253-315)
MGNSSSEPKPPTKSDKPSPPVDQTNVHVYPDWAAMQAYYGPVAMPYPYNSAMAASGHP
PPPYMWNPHMMSPSGAPYAAVYPHGGGVYAHGPIPMGSLPQGQKDFPLTTPGTLLSIDT
PTKSTGNTDNGLMKKLKEFDGLAMSLGNGNPENGADCHKRSRNSSETDGSIDGSDGNTTG
ADEPKLKRSTREGTPTKDGKQLVQASSFHSVSPSSGDTGVKLIQSGGAILSPGVNSANSNPF
MSQSLAMVPPETWLNRELERKRRERKQSNRESARRSRLRKQAEETEELARKVEALTAENMA
LRSELNQLNEKSKLRGANATLLDKLKCSEPEKRVANMLSRVKNAGAGDKNKNQGDNDNS
NSTSKFHQLLDTKPRAKAVAAG*

>G591 (88..1020)

GTAAATCTCTCTTTGAAGGTTTCCTAACTCGTTAATCGTAACCTCACAGTGACTCGTTTCGAG
TCAAAGTCTCTGTCTTTAGCTCAAACCATGGCTAGTAACAACCTTCACGACAACCTTTCT
GACCAAACCTCTTCTGATGATTTCTTCGAGCAAATCCTCGGCCTTCTAACTTCTCAGCC
TCTTCTGCGCGCGGTTTATCTGGAGTTGACGGAGGATTAGGTGGTGGAGCACCCTATG
ATGCTGCAGTTGGGTTCCGGAGAGAAGGAAGTACATGGGTGGCTTAGGAGGAAGTGGA
CCAACTGGGTTTACAATCAGATGTTTCTTTGGGGTTAAGTCTTGATCAAGGGAAAGGA
CCTGGGTTTCTTAGACCTGAAGGAGACATGGAAGTGGGAAAAGATTCTCAGATGATGTT
GTTGATAATCGATGTTCTTCTATGAAACCTGTTTCCACGGGCAGCCTATGCAACAGCCA
CCTCCATCGGCCCCACATCAGCCTACTTCAATCCGTCCTCCAGGGTTCGAGCTAGGCGTGGT
CAGGCTACTGATCCACATAGCATCGCTGAGCGGCTACGTAGAGAAAGAATAGCAGAACGG
ATCAGGGCGCTGCAGGAACCTGTACCTACTGTGAACAAGACCGATAGAGCTGCTATGATC
GATGAGATTGTCGATTATGTAAGTTTCTCAGGCTCCAAGTCAAGGTTTGGAGCATGAAC
CGACTTGGTGGAGCCGGTGGGTTGCTCCACTTGTACTGATATGCCTCTTTCATCATCA
GTTGAGGATGAAACGGGTGAGGGTGGAAGGACTCCGCAACCAGCGTGGGAGAAATGGTCT
AACGATGGGACTGAACGTCAGTGGCTAACTGATGGAAGAGAACGTTGGAGCCGCGATG
CAGCTTCTTCAATCAAAGGCTCTTTGTATGATGCCAATCTCATTTGGCAATGGCAATTTAC
CATCTCAACCTCCGGATACATCTTCAGTGGTCAAGCCTGAGAACAATCTCCACAGTAG
GATTTCTGCAATAAAGAGTTTGTACAGCTAATCCAACGTCCAACATGGGTTTTTCTTCT
GCTCTAATGACTCTGGTTTTCTTCTCCTCTCTCACCAGCTTGAAAGGTAAAAAAGTGAA
AAAGGCTTTGTAGATGGAATCAATGTAGGATTGTCAGTAGAGGGCAAAAAAATGTCATAT
AGCTCAATTGATCAAGTCTTAAAAAATAAAAAA

>G591 Amino Acid Sequence (domain in AA coordinates: 143-240)
MASNNPHDNLSDQTPSDDFFEQILGLPNFSASSAAGLSGVDGGLGGGAPPMMLQLGSGBE
GSHMGLGSGSGPTGFHNQMFPLGLSLDQKGPGFLRPEGGHSGSKRFSDVDVNDNRCSMK
PVFHGQPMQPPPSAPHQPTSIRPRVRARRGQATDPHSIAERLRERRERIAERIRALQELVP
TVNKTDRAMIDEIVDYVKFLRLQVKVLSMNLGGAGAVAPLVDMPLSSSVEDETGE
RTPQPAWEKWSNDGTERQVAKLMEENVGAAMQLLQSKALCMMPISLAMAIYHSQPPDTSS
VVKPENPPQ* -

>G8 (247..1596)

AAAAAAAATATCCGTCTCACTCTCTCGCCCGGTAACATTTCCCGCGACAAACTTC
TCTACTCTCACCATTCTCCATCGTAATCTCTAAATCTTCTCATTCTCTTCTCTCC
CGATCATCTCGAGCTCTCGTGAGAGATTATGTGATTATGTAATCGTTGTTGCTGTAGAA
GACGATCTCTAAACAACCTTCTCTCATCATCACCTTCGCTAGATTGTAATTTTCAGAG
CTTGAGATGTTGGATCTTAACCTCAACGCTGATTCTCCCGAGTCGACTCAGTACGGTGGT
GACTCATACTTAGATCGGCAGACATCAGACAACTCCGCCGGGAATCGAGTGAAGAGTCC
GGTACATCGACGTCGTAGTTATCAATGCCGATGGAGACGAAGACTCTTGCTCTACTCGA
GCTTTCACTCTCAGTTTCGATATTTTAAAGTCGGAAGTAGTAGCGCGGAGACGAAAGC

CCGCGCGCTTCAGCTTCCGTTACTAAAGAGTTTTTTCCGGTGAGTGAGAGACTGTGGACATCTACGAGATGTTGAAGGATCATCAAGCTCTAGAACTGGATAGATCTTTCTTTTGACCGTATTGGTGACGGAGAAACGAAATTGGTAACTCCGGTTCGACTCCGGCTCCGGTTCGGGCTCAGGTTAAAAAGAGTCCGAGAGGACCAAGGTCTAGAAGTTCACAGTATAGAGGAGTTACTTTTATAGAAGAAGCTGGTCGATGGGAGTACATATTTGGGATTGTGGGAAACAGTTTTATTAGGTGGTTTCGACTGCTCATGCTGCAGCTAGAGACTATGATCGAGCTGCTATTAAATTTAGAGGTTGTGATGCTGATATCAACTTTACTCTTGGTGATTATGAGGAAGATATGAAACAGGTACAAAAGTTAGTAAGGAAGAGTTTTGTGCATATACTCGCTAGACAGACGCGGGTTTTTCGCGGGGAGTTTCAAGTATCGAGGGGTTACGTTACACAAATGTGGTAGATGCGGAACTAGGATGGGGCAGTTCTTGGTAAAAAGGCTTATGACAAGGCTGCAATCAACACTAATGGTAGAGAAGCAGTCACGAAGCTTCGAGATGAGTTATACCAAAATGAGATTAACTCTGAGAGCAATAACTCTGAGATTGACCTCAACTTGGGAATCTCTTTATCGACCGGTAATGCGCCAAGCAAAATGGGAGGCTCTTTCACCTCTCCCTCTAATACTTATGAAACTCAGCGTGAGAGTTAGCTTGAGGATAGATAACGAATACATGGGAAAGCCGGTGAATACACCTCTCTCTTATGGAATCCTCGGATCATCGCCTTTACTGGAACCGAGCATGCCCCGAGTTATAATAATCCCGCCGAGGAAGAGCAACAGAAAAGAGAAGTGAAGCTGAAGGGATGATGAGTAAGTGGGGATGGCAGAGACCGGGGCAACAGCGCCGTGAGACCGCAGCCACCGGGACCACAACCACCACCATTTCTCTCAGTTGCAGCAGCATCATCAGGATTCTCAGATTTCCGGCCACAACCTCCCAATGACAATGCAACAGCTGTTACTTTTATCCACACCCCTTAACTTGTAGGGGACATATGAGAGTTTTTTTACCATCTCTCTCTCTCAACACTCTAGTCCCCCTTTCAAAATGTCATTTGGGTTTAGATTTTTTACATACAATGATCAATTTTTCC

>G8 Amino Acid Sequence (domain in AA coordinates: 151-217, 243-296)

MLDLNLNADSPSTQYGGDSYLDLDRQTSNDSAGNRVEESGTSTSSVINADGDEDSCTSTRAF
TLSFDILKVGSSSGGDESPAASASVTKFFPVSGDCGHLRDVEGSSSRNWIDLSFDRIG
DGETKLVTVPPTAPVPAQVKSRRGPRSSSQYRGVTFYRRTRGWESHIWDCGKQVYLG
GFDTAHAARAYDRAAIKFRGVADINFTLGDYEEEMKQVQNLKSKEEFVHILRRQSTGFS
RGSSKRYRGVTLHKCGRWEARMGQFLGKKAYDKAAINTNGREAVTNFEMSSYQNEINSESN
NSEIDLNLGISLSTGNAPKONGRLFHFPSEINTYETQRGVSLRIDNEYMGKPVNTPLPYGSS
DHRLYWNGACPSYNNPABEGRATEKRSEAEGMMSNWGWQRPQGTSAVRPQPPGPQPPPLFS
VAAASSGFSHFRPQPPNDNATRGYFYPHP*

>G859 (162..752)

GATTTGTCAATTTTGTCTAGCCAAAAAAAAAAAAAAAAAGGAGAGAGAGAGAGAGA
GAGAGAGAGAGAAACGAAGAAAAAAAAAAGCAAAAAACATTGTGGGTCTCCGGTGATT
AGGATCAAATTAGGGCACCAGCCTTATCGGAGGAAGAAGCCATGGGTAGAAAAAAGTCG
AGATCAAGCGAATCGAGAAACAAAAGTAGTCGACAAGTCACTTTCTCCAAACGACGCAATG
GTCTCATCGAGAAAGCTCGACAACCTTTCAATTCTCTGTGAATCTTCCATCGCTGTTCTCG
TCGTCTCCGGTCTCCGGAAAACTCTACAAGTCTGCGCTGCGGTGACAACATGTCAAAAGATCA
TTGATCGTTACGAAATACATCATGCTGATGAACCTTGAAGCCTTAGACTCTTGCGAGAAAAA
CTCGGAATTATCTGCCACTCAAAGAGTTACTAGAAATAGTCCAAAGCAAGCTTGAAGAAT
CAAATGTCGATAATGCAAGTGTGGATACTTTAATTTCTCTGGAGGAACAGCTCGAGACTG
CTCTGTCCGTAACTAGAGCTAGGAAGACAGAACTAATGATGGGGGAAGTGAAGTCCCTTC
AAAAACCGGAGAACTTGCTGAGAGAGAGAGAACAGACTTTGGCTAGCCAGGTGGGGGAAGA
AGACGTTTCTGGTTATAGAAGGTGACAGAGGAATGTCAATGGGAAATGGCTCCGGCAACA
AAGTACGGGAGACTCTCCGCTGCTCAAGTAATCACCATCATCAACGGCTGAGCTTTCAC
CTTAACTTACAGCCTGATTCAGAAGTTTTTACAAATTTGTAAATTATAAAAAAGCTTTCAT
AATAATCTCAACCTTTTATCTTCTCGCGCCAATGTGGAAATTAAGGTTAAAAATAAAA
TAAAAACAGAAGCTCATCGGAAAGAAATTGTAAAACCTAAGATAAAGCTATAGTAGATCTTTA
TTGTACCTTCTGATACGATAAAGATTTATTCTGTGTGTTTGTCTTCCCTCNAAAAAAAAA
AAAAAAAAAAAAAAAAAA

>G859 Amino Acid Sequence (domain in AA coordinates: TBD)

MGRRKKEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVSGSGKLYKSASG
DNMSKIIDRYEIIHHADELEALDLAEKTRNYLPLKELLEIVQSKLEESNVDNASVDTLISL
EEQLETALSVTRARKTELMMEVKSLSQKTENLLREENQTLASQVGKKTFLVIEGDRGMSW
ENGSGNKVRETLPLLK*

>G878 (197..1738)

CAAAAAAATCTCTCCATTAAAAGACTGCCCAAAGAAATATTTTATACAAAATGAAAGA
GAGAAACACGACACGAATTTTGTATAATTAAGATTACACAAAAAAGTGTTAGAAAGAG

AAATATCTTCTCTTTTCTGTGTGAGTTGGGTTTGTAAAGTTTTATCCTTTTGTTC
TCAAAATCAAGAATCGATGGCGGAGAAGGAAGAAAAAGAACCATCGAAGTTAAATCATC
CACCGGAGTTTACGGCCAACGATTTCACTACCTCCTCGACCGTTTGGTGAAATGTTTTT
TAGCGGTGGCGTTGGATTTAGTCCTGGACCAATGACTCTCGTCTCAAATTTATTCTCTGA
TCCTGATGAGTTCAAGTCTTTCTCTCAGCTTTTAGCTGGAGCTATGGCTTCTCGGCGGC
AGCTGCTGTTGCCGCGCTGCTGTGGTTGCTACTGCTCATCATCAGACACCTGTGAGCTC
TGTCGGTGATGGCGGTGGAAGCGGTGGTGATGTTGACCCGAGGTTTAAAGCAGAGTAGACC
AACGGGATTGATGATAACTCAACCACCGGGGATGTTTACTGTACCGCCGGGGTTAAGTCC
GGCTACTCTTTTGGATTTCTCCGAGCTTCTTTGGTCTTTTTCACCTCTTCAGGGAACATT
TGGTATGACACATCAACAAGCTTTAGCACAAAGTCACTGCACAAGCAGTTCAAGGCAATAA
TGTTTCATATGCAGCAATCAACAATCTGAATATCCTTCTTCTACACAACAACAACA
ACAACAACAAGCTTTCATTGACTGAGATTCCATCATTTTCTTCTGCACCTAGGTCTCA
GATTCGAGCCTCGGTTCAGAAACATCGCAGGGTCAGAGAGAGACTTCGGAAATATCTGT
CTTTGAGCATCGGTTCACAGCCTCAAAATGCTGACAAACCAGCTGATGATGGATACAACCTG
GCGGAAATATGGGCAGAAGCAAGTGAAGGGGAGCGATTTTCTCGGAGTTATTACAAATG
TACGCATCCAGCTTGTCTGTCAAGAAGAAAGTGGAGAGGTCACTCGATGGACAAGTAAC
GGAAATCATCTACAAGGGTCAACACAATCATGAGCTTCTTCAAAGCGCGGTAACAATAA
CGGGAGTTGTAAAGTTCTGATATTGCAAATCAGTTTCAAACAAGTAATAGCAGTCTCAA
CAAGAGTAAGAGGGACCAGGAAACAAGCCAAGTTACAACAACAGAGCAGATGTCTGAAGC
AAGTGATAGCGAGGAGGTTGGGAATGCAGAGACTAGTGTGGGAGAAAGACATGAGGATGA
GCCTGATCCCAAGCGAAGAAATACAGAAGTTCCGGTTTCAGAACCAAGTTGCTTCATCGCA
TAGAACTGTGACAGAGCCTAGGATTATTGTCCAAACGACGAGTGAAGTTGACCTCTTAGA
TGATGGATATAGTTGGCGCAAGTATGGTCAGAAAGTAGTCAAAGGAAATCCTTATCCGAG
GAGCTACTATAAGTGATGTAACAACAGATTGCGGAGTAAGGAAACATGTAGAGAGAGCAGC
AACTGACCCAAAAGCTGTTGTAACAACATATGAAGGTAAACATAACCATGATGTTCCAGC
TGCTAGAACCAGCAGCCATCAGTTAAGACCAACAATCAACACAACACCTCAACGGTTAA
CTTCAATCATCAACAGCCTGTTGCACGTTTAAAGCTTAAAGAAGAGCAAATCACTTGACA
GAGAAGAAGAAATACGACGGCGCTTGAGCTTTTGTGAGTTTAAATGAATCTTCTTTTGGTT
AATGAACCTGTTTTTGTGCTCAAAACACCACAGGTTTCTCTGGACAGAATCTCTGATA
TTACAGTTTCAAAGGTATGTTCTTTTATTTTCATGTTGGAATCTTCTGTGTAATCTTAAG
AAGCTTTAGGAGGTAAATGTAACAAAACAGATTCAAAGTTATGCCCTTATGTGAATCTTT
TGACATGGGATAAACAAAATTTACAGGTATCCTTTTGTCTTGTGTAAAAA

>G878 Amino Acid Sequence (domain in AA coordinates:250-305, 415-475)

MAEKEEKEPSKLKSSSTGVSRTISLPPRPFGEFFSGGVGFSPGPMTLVSNLFSDFDEFK
SFSQLLAGAMASPAAAVAAAVVATAHHQTPVSSVGDGGSGGDVDPRFKQSRPTGLMI
TQPPGMFTVPPGLSPATLLDSPSFFGLFSPLOQTFGMTHQQALAQVTAQAVQGNVHMQQ
SQQSEYPSSTQQQQQQQOQSLTEIPSFSSAPRSQIRASVQETSQOQRETSEISVFEHRS
QPQNAADKPADDGYNWRKYQKQVKGSDFPRSYKCTHPACPVKKKVERSLDGQVTEIIYK
GQHNHELPQKRGNNNGSCKSSDIANQFQTSNSSLNKSQRDQETSQVTTTEQMSEASDSEE
VGNAETSVGERHEDEPDPRNRTEVRVSEPVASSHRTVTEPRIIVQTTSEVDLLDDGYRW
RKYGQKVVKGPNPYPRSYKCTTPDCGVRKHVERAATDPKAVVTTYEGKHNHDVPAARTSS
HQLRPNNQHNTSTVNFNHQOPVARLRLKEEQIT*

>G971 (131..1171)

TTTTTTTCTTCCCTCTTTTAGAAGTCTCTCTCTCTCTCGTTTTTGACACTTATCCTCTC
TCTTTTTTCTCTCTCCCTCTCTCTCTGGCCGGAAGAAACACGTCGTTTATAGCTAA
AGATTCGATCATGTTGGATCTTAACCTAAAGATCTTTTCTTCTTATAACGAAGATCAAGA
TCGGAAAGTACCATTAAATGATCTCAACCACCGGTGAAGAAGAATCTAACTCATCTTCCTC
CTCCACAACAGACTCTGCAGCGAGAGATGCTTTTCATCGCTTTTGGAAATCTCAAACGCGA
CGATGACCTTGTTCCTCCTCCTCCTCCTCCTCCTCATAAGAAACAGGAGATCTCTTCC
GGTGGTGGCTGATGCTCGTGGGAATATAGAATCTCCGTGGAAGACAGTCACTGGTTGAA
TCTTTCTCTTTTACAAGAAATACACAGAAATGGTGAAGAAGAGCAGAAGAGGACCAAG
GTCTCGTAGCTCCCAATATCGTGGCGTCACTTTTTTACCGTCCGACCGGTCTGTTGGGAATC
TCATATTTGGGATTGTGGAAGCAAGTTTATTTGGGCGGGTTGATACTGCTTACGCAGC
AGCAAGGGCTTACGACCGAGCTGCTATCAAATTCGGTGGTCTCGATGCAGACATCAATTT
CGTCTGGATGATTATAGGCATGACATCGATAAGATGAAGAATTTAAATAAGGTGGAGTT
CGTGCAAAACACTTAGGCGAGAGAGTGCGAGTTTTCGGAAGAGGAAGTTCCAAATACAAAGG

CTTGGCTCTTCAAAAATGCACCCAATTCAAAACTCATGATCAGATTCATCTCTTCCAAAA
CAGGGGATGGGATGCAGCAGCAATAAAATACAATGAGTTGGGAAAGGGAGAAGGAGCCAT
GAAGTTTGGTGCCCATATCAAAGGAAATGGTCACAATGATCTTGAACCTAGTCTCGGAAT
TTCATCATCATCGGAAAGTATAAAGTTGACAACAGGCGATTACTATAAGGGTATCAATCG
GTCCACGATGGGTTTATACGGTAAGCAATCATCGATATTTTACCCATGGCAACCATGAA
ACCTCTGAAGACAGTTGACGATCATCAGGATTCCTTTTATCAGCATGACAAGTTCCTC
TTCTCCATGTCCAATTGTTTGTATCCATAGGATCGTTCTACACTCTCTTAACCTAATATA
TATTTTTACTCTATCTGATTATTGTATACAAGGATAAAATTTGATTCTTTCCTTAATGAG
TGAGAAATATTGGAAGTGTTAAAAA

>G971 Amino Acid Sequence (conserved domain in aa coordinates: 120-186)

MLDLNLKIFSSYNEDQDRKVPMLMISTTGEESNSSSSSTDSAADAFIAFGILKRDDDL
VPPPPPPPHKETGDLFPVADARNNIEFVSDSHWLNLSLQNRNTQKMKVKSRRGPRSR
SQYRGVTFYRRTGRWESHIWDCGKQVYLGFDYAAAAAYDRAAIKFRGLDADINPVVD
DYRHDIDKMKNLNKVEFVQTLRRESASFGRSSSKYKGLALQKCTQFKTHDQIHLFQNRGW
DAAAIKYNELGKGEGAMKFGAHIKNGHNDLELSLGISSSESILKTTGDYKGINRSTM
GLYKQSSIPLPMATMKPLKTVASSGFPFISMTSSSSSMSNCFDP*

>G975 (58..657)

ATTACTCATCATCAAGTTCTCTCTCTGACAAACATCACAGAGTAAGTAAGAATG
GTACAGACGAAGAAGTTCAGAGGTGTGAGGCAACGCCATTGGGGTCTTGGGTCGCTGAG
ATTCGTCTCTCTCTTGAACGGAGGATTGGCTAGGGACGTTTCGAGACCGCAGAGGAG
GCAGCAAGACATACGACGAGGCCCGCTTTAATGAGCGCCGCAACGCCAAAACCAAC
TTTCCCCTCAACAACAACAACACCGGAGAACTTCCGAGGGCAAAACCGATATTTCAGCT
TCGTCCACAATGTCTCTCAACATCATCTTCATCGCTCTCTTCCATCCTCAGCGCCAAA
CTGAGGAAATGCTGCAAGTCTCTTCCCCTACCTGCCTCCGTCTTGACACAGCC
AGCTCCCATATCGGCGTCTGGCAGAAACGGGCGGTTCAAAGTCTGACTCCAGCTGGGTC
ATGACGGTGGAGCTAGGTCCCGAAGCTCCTCCAAGAGACTACTAGTAAAGCTTCACAA
GACGCTATTCTTGCTCCGACCACTGAAGTTGAAATTGGTGGCAGCAGAGAAGAAGTATTG
GATGAGGAAGAAAAGGTTGCTTTGCAAATGATAGAGGAGCTTCTCAATACAAACTAAATC
TTATTTGCTTATATATATGTACCTATTTTCATTGCTGATTTACAGCCAAAATAATCAATT
ATACCGTGATTTTATAGATGTTTTATATTAAAGGTTGTTAGATATA

>G975 Amino Acid Sequence (domain in AA coordinates: 4-71)

MVQTKKFRGVRQRHWGWSVAEIRHPLLRRIWLGTFTAEAAARAYDEAAVLMSGRNAKT
NFPLNNNTGETSEGKTDISASMTSSSTSSSSLSILSAKLKCKCKSPSPSLTCLRLDT
ASSHIGVWQKRAGSKDSSWMTVELGPASSSQETTSKASQDAILAPTTEVEIGGSREEV
LDEEEKVALQMIEELLNTN*

>G994 (180..917)

TGTATATATAGTTAGTTAGTTGAGATAAACTTGGTTACCACTTTTGTGTGGTCTTTCTTT
TTCTTTTCTCCATTTTCCATTTATCGACCCCTTGGGTGTAGCTAATTACTTTTCGCGATT
TTCAAATCCAATAAAGTTTAAATTTGATGAAGCTTTTAAACCATATAATATAAATAA
TGGGTGGTTCGTAAACCATGTTGTGATGAGGTTGGATTAAGAAAGGGTCCATGGACAGTGG
AAGAAGATGGGAAACTAGTTGATTTCTTAAGGGCACGTGGCAACTGCGGTGGTGGTGGAG
GAGGATGGTGTGGAGAGAGCGTGCCAAAACCTGGCGGGGCTAAGGAGGTGTGGCAAAAGTT
GCCGTCTCCGGTGGACTAATTATCTCCGGCCAGATCTCAAGAGAGGTCTTTTACTGAAG
AAGAAATCCAACCTAGTCATTGATCTTCATGCTCGCCTTGGCAATAGATGGTTCGAAGATTG
CAGTGGAGTTACCAGGAAGAACAGACAACGATATCAAAAATTATTGGAACACTCATATAA
AGAGGAAGCTTATAAGAAATGGGTATTGATCCAAACACACATCGTCGATTTGACCAACAAA
AAGTCAACGAGGAGGAAACGATATTGGTCAACGATCCAAAGCCTCTGTCTGAGACCGAGG
TATCTGTGCTTTGAAGAATGACACGTGAGAGTGTATCAGGAAATCTAAACCAATTGG
CTGACGTGGACGGTGATGATCAGCGGTGGAGCTTTCTAATGGAAAATGACGAAGGAGGAG
GTGGCGACGCCCGGAGAGCTTACGATGCTATTGTCCGGTGACATTACGTCATCATGTT
CTTCTTCGTCTCTTTGTGGATGAAGTATGGAGAATTTCGGATACGAAGATTTAGAAGTTG
GATGTTTCGATGTTTGAAGATTCAAGTATGTTTAAATTAGGCCGTAGGTTGATTAATCATA
AGGTTTCATTGACTTCATTCTAGAATTGTGTAGTTGGACCAGTATAAGAATCAAAGTTAT
GAAACATTGTAATTTGATTTCCAAATTAATCTAATGAATAAATGTGCTTTGCAAAAAA
AAAAA

>G994 Amino Acid Sequence (domain in AA coordinates: 14-123)
MGGRKPCDEVGLRKGPWTVEEDGKLVDLFRARGNCGGGGGWCWRDVPKLAGLRRCGKS

CRLRWNTNYLRPDLKRGLFTEEEIQLVIDLHARLGNRWSKI AVELPGRDNDIKNYWNTHI
KRKLIRMGIDPNTHRRFDQQKVNEEETILVNDPKPLSETEVSVALKNDTS AVLSGNLNQL
ADV DGGDQ PWSFLMENDEGGGGDAAGELTMLLSGDITSSCSSSSSLWMKYGEFGYEDLEL
GCFDV*

>G2347 (81..626)

AGCCCATCCTTCAACATTGCTTCCTAACCAGAAATCCACCATCATCTTCCACGAATACA
ACTTAAAGCTTTACAGAAAATGGAGGGTCAGAGAACAACGCCGGGGTTACTTGAAAG
ACAAGGCTACAGTCTCCAACCTTGTGGAAGAAGAAATGGAGAATGGCATGGATGGAGAAG
AGGAGGATGGAGGAGACGAAGACAAAAGGAAGAAGGTGATGGAAAGAGTTAGAGGTCCTA
GCACTGACCGTGTTCCATCGCGACTGTGCCAGGTCGATAGGTGCACTGTTAATTTGACTG
AGGCCAAGCAGTATTACCCGACAGACAGAGTATGTGAAGTACATGCAAAGGCATCTGCTG
CGACTGTTGCAGGGGTCAGGCAACGCTTTTGTCAACAATGCAGCAGGTTTCATGAGCTAC
CAGAGTTTGTGAAGCTAAAAGAAGCTGCAGGAGGCGCTTAGCTGGACACAATGAGAGGA
GGAGGAAGATCTCTGGTGACAGTTTTGGAGAAGGGTCAGGCCGGAGAGGGTTTAGCGGTC
AACTGATCCAGACTCAAGAAAGAAACAGGGTAGACAGGAAACTTCCTATGACCAACTCAT
CATTCAAGCGACCACAGATCAGATAAACCTCCCGCTCTCTCTCTTCTGTCTATCTACATA
TGCTCTATCTACTCTTATTAGACAAATAATGGCATCTAACAATGTCAAGAAAAGTTGG
TCATGGTATTAAATCCTACCGGATATATACTATAAACTCTAGTCCCTCTATGCTGT
CCTGTAATGAATATCTATCCGGAATGTATTGCGCATAGTCTTGCGTCTAATAATGTTTAT
TGATTTTGTA

>G2347 Amino Acid Sequence (domain in AA coordinates: 60-136)
MEGQRTQRRGYLKDKATVSNLVEEEMENGMDGEEEDGGDEDKRKKVMERVGRGPSTDRVPS
RLCQVDRCTVNLTEAKQYRRHRVCEVHAKASAATVAGVRQRFCCQCSRFFHELPEFDEAK
RSCRRRLAGHNERRRKISGDSFGEESGRRGFSGQLIQTQERNRVDRKLPMTNSSFKRPQI
R*

>G2010 (1..525)

ATGGAGGGTAAGAGATCACAAGGACAAGGTTACATGAAAAAGAAGTCTTACCTTGTGGAA
GAAGATATGGAGACTGATACGGATGAAGAAGAGGAAGTAGGTAGGGATAGAGTTAGAGGG
TCTAGAGGTAGCATCAATCGTGGTGGCTCGTTGCGGCTTTGCCAAGTAGATAGATGCACA
GCTGATATGAAAGAGGCAAACTGTATCACCAGGAGACACAAAGTGTGTGAAGTTCATGCA
AAGGCATCTTCTGTCTTTCTCTCAGGACTTAACCAACGCTTTTGTCAACAATGCAGTAGG
TTTCATGACCTCCAAGAGTTTGATGAAGCTAAGAGAAGTTGCAGGAGGCGCTTAGCTGGA
CACAATGAGCGAAGAAGGAAGAGCTCTGGTGAGAGTACTTATGGAGAAGGATCAGGTCGG
AGAGGAATCAATGGTCAGGTGGTGTATGCAGAATCAAGAAAGATCAAGGGTAGAGATGACA
CTTCCTATGCCAAACTCATATTCAAGCGACCACAGATTAGATAG

>G2010 Amino Acid Sequence (domain in AA coordinates: 53-127)
MEGKRSQGGQYMKKKSYLVEEDMETDTDEEEVGRDRVRGSRGSINRGGSLRLCQVDRCT
ADMKEAKLYHRRHKVCEVHAKASSVFLSGLNQRFCCQCSRFDLQEFDEAKRSCRRRLAG
HNERRRKSSGESTYGEESGRRGINGQVVMQNQERSRVEMTLMPNPSSFKRPQIR*

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
20 February 2003 (20.02.2003)

PCT

(10) International Publication Number
WO 03/013227 A2

(51) International Patent Classification⁷: A01H

(US). YU, Guo-Liang [US/US]; 242 Gravett Drive, Berkeley, CA 94705-1531 (US). BROUN, Pierre, E. [FR/US]; 921 Sunnybrae Boulevard, San Mateo, CA 94402 (US).

(21) International Application Number: PCT/US02/25805

(22) International Filing Date: 9 August 2002 (09.08.2002)

(74) Agents: WARD, Michael, R. et al.; Morrison & Foerster LLP, 425 Market Street, San Francisco, CA 94105-2482 (US).

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/310,847	9 August 2001 (09.08.2001)	US
60/336,049	19 November 2001 (19.11.2001)	US
60/338,692	11 December 2001 (11.12.2001)	US
10/171,468	14 June 2002 (14.06.2002)	US

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(71) Applicant (*for all designated States except US*): MENDEL BIOTECHNOLOGY, INC. [US/US]; 21375 Cabot Boulevard, Hayward, CA 94545 (US).

(84) Designated States (*regional*): ARIPO patent (GI, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): RATCLIFFE, Oliver [GB/US]; 814 East 21st Street, Oakland, CA 94606 (US). RIECHMANN, Jose, Luis [ES/US]; 518 S. El Molino Avenue, #308, Pasadena, CA 91101 (US). ADAM, Luc, J. [CA/US]; 25800 Industrial Boulevard, Apt. L403, Hayward, CA 94545 (US). DUBELL, Arnold, N. [US/US]; 14857 Wake Avenue, San Leandro, CA 94578 (US). HEARD, Jacqueline, E. [US/US]; 810 Guilford Avenue, San Mateo, CA 94402 (US). PILGRIM, Marsha, L. [US/US]; 1368 Patrick Henry Drive, Phoenixville, PA 19460 (US). JIANG, Cai-Zhong [US/US]; 34495 Heathrow Terrace, Fremont, CA 94555 (US). REUBER, T., Lynne [US/US]; 1115 S. Grant Street, San Mateo, CA 94402 (US). CREELMAN, Robert, A. [US/US]; 2801 Jennifer Drive, Castro Valley, CA 94546 (US). PINEDA, Omaira [CO/US]; 4060 9th Place, Vero Beach, FL 32960

Published:

— without international search report and to be republished upon receipt of that report

(48) Date of publication of this corrected version:

3 July 2003

(15) Information about Correction:

see PCT Gazette No. 27/2003 of 3 July 2003, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: YIELD-RELATED POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS

(57) Abstract: The invention relates to plant transcription factor polypeptides, polynucleotides that encode them, homologs from a variety of plant species, and methods of using the polynucleotides and polypeptides to produce transgenic plants having advantageous properties compared to a reference plant. Sequence information related to these polynucleotides and polypeptides can also be used in bioinformatic search methods and is also disclosed.

WO 03/013227 A2

YIELD-RELATED POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS

This application claims the benefit of US Provisional Application No. 60/310,847, filed August 9, 2001, US Provisional Application No. 60/336,049, filed December 5, 2001, US Provisional Application No. 60/338,692, filed December 11, 2001, and US Non-provisional Application No. 10/171,468, filed June 14, 2002, the entire contents of which are hereby incorporated by reference.

FIELD OF THE INVENTION

This invention relates to the field of plant biology. More particularly, the present invention pertains to compositions and methods for phenotypically modifying a plant.

INTRODUCTION

A plant's traits, such as its biochemical, developmental, or phenotypic characteristics, may be controlled through a number of cellular processes. One important way to manipulate that control is through transcription factors - proteins that influence the expression of a particular gene or sets of genes. Transformed and transgenic plants that comprise cells having altered levels of at least one selected transcription factor, for example, possess advantageous or desirable traits. Strategies for manipulating traits by altering a plant cell's transcription factor content can therefore result in plants and crops with commercially valuable properties. Applicants have identified polynucleotides encoding transcription factors, developed numerous transgenic plants using these polynucleotides, and have analyzed the plants for a variety of important traits. In so doing, applicants have identified important polynucleotide and polypeptide sequences for producing commercially valuable plants and crops as well as the methods for making them and using them. Other aspects and embodiments of the invention are described below and can be derived from the teachings of this disclosure as a whole.

BACKGROUND OF THE INVENTION

Transcription factors (TFs) can modulate gene expression, either increasing or decreasing (inducing or repressing) the rate of transcription. This modulation results in differential levels of gene expression at various developmental stages, in different

tissues and cell types, and in response to different exogenous (e.g., environmental) and endogenous stimuli throughout the life cycle of the organism.

Because transcription factors are key controlling elements of biological pathways, altering the expression levels of one or more transcription factors can change entire biological pathways in an organism. For example, manipulation of the levels of selected transcription factors may result in increased expression of economically useful proteins or metabolic chemicals in plants or to improve other agriculturally relevant characteristics. Conversely, blocked or reduced expression of a transcription factor may reduce biosynthesis of unwanted compounds or remove an undesirable trait. Therefore, manipulating transcription factor levels in a plant offers tremendous potential in agricultural biotechnology for modifying a plant's traits.

The present invention provides novel transcription factors useful for modifying a plant's phenotype in desirable ways.

SUMMARY OF THE INVENTION

In a first aspect, the invention relates to a recombinant polynucleotide comprising a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence encoding a polypeptide comprising a polypeptide sequence selected from those of the Sequence Listing, SEQ ID NOs:2 to 2N, where $N = 2-561$, or those listed in Table 4, or a complementary nucleotide sequence thereof; (b) a nucleotide sequence encoding a polypeptide comprising a variant of a polypeptide of (a) having one or more, or between 1 and about 5, or between 1 and about 10, or between 1 and about 30, conservative amino acid substitutions; (c) a nucleotide sequence comprising a sequence selected from those of SEQ ID NOs:1 to $(2N - 1)$, where $N = 2-561$, or those included in Table 4, or a complementary nucleotide sequence thereof; (d) a nucleotide sequence comprising silent substitutions in a nucleotide sequence of (c); (e) a nucleotide sequence which hybridizes under stringent conditions over substantially the entire length of a nucleotide sequence of one or more of: (a), (b), (c), or (d); (f) a nucleotide sequence comprising at least 10 or 15, or at least about 20, or at least about 30 consecutive nucleotides of a sequence of any of (a)-(e), or at least 10 or 15, or at least about 20, or at least about 30 consecutive nucleotides outside of a region encoding a conserved domain of any of (a)-(e); (g) a

nucleotide sequence comprising a subsequence or fragment of any of (a)-(f), which subsequence or fragment encodes a polypeptide having a biological activity that modifies a plant's characteristic, functions as a transcription factor, or alters the level of transcription of a gene or transgene in a cell; (h) a nucleotide sequence having at least 31% sequence identity to a nucleotide sequence of any of (a)-(g); (i) a nucleotide sequence having at least 60%, or at least 70 %, or at least 80 %, or at least 90 %, or at least 95 % sequence identity to a nucleotide sequence of any of (a)-(g) or a 10 or 15 nucleotide, or at least about 20, or at least about 30 nucleotide region of a sequence of (a)-(g) that is outside of a region encoding a conserved domain; (j) a nucleotide sequence that encodes a polypeptide having at least 31% sequence identity to a polypeptide listed in Table 4, or the Sequence Listing; (k) a nucleotide sequence which encodes a polypeptide having at least 60%, or at least 70 %, or at least 80%, or at least 90 %, or at least 95 % sequence identity to a polypeptide listed in Table 4, or the Sequence Listing; and (l) a nucleotide sequence that encodes a conserved domain of a polypeptide having at least 85%, or at least 90%, or at least 95%, or at least 98% sequence identity to a conserved domain of a polypeptide listed in Table 4, or the Sequence Listing. The recombinant polynucleotide may further comprise a constitutive, inducible, or tissue-specific promoter operably linked to the nucleotide sequence. The invention also relates to compositions comprising at least two of the above-described polynucleotides.

In a second aspect, the invention comprises an isolated or recombinant polypeptide comprising a subsequence of at least about 10, or at least about 15, or at least about 20, or at least about 30 contiguous amino acids encoded by the recombinant or isolated polynucleotide described above, or comprising a subsequence of at least about 8, or at least about 12, or at least about 15, or at least about 20, or at least about 30 contiguous amino acids outside a conserved domain.

In a third aspect, the invention comprises an isolated or recombinant polynucleotide that encodes a polypeptide that is a paralog of the isolated polypeptide described above. In one aspect, the invention is an paralog which, when expressed in *Arabidopsis*, modifies a trait of the *Arabidopsis* plant.

In a fourth aspect, the invention comprises an isolated or recombinant polynucleotide that encodes a polypeptide that is an ortholog of the isolated polypeptide described above. In one aspect, the invention is an ortholog which, when expressed in *Arabidopsis*, modifies a trait of the *Arabidopsis* plant.

In a fifth aspect, the invention comprises an isolated polypeptide that is a paralog of the isolated polypeptide described above. In one aspect, the invention is an paralog which, when expressed in *Arabidopsis*, modifies a trait of the *Arabidopsis* plant.

In a sixth aspect, the invention comprises an isolated polypeptide that is an ortholog of the isolated polypeptide described above. In one aspect, the invention is an ortholog which, when expressed in *Arabidopsis*, modifies a trait of the *Arabidopsis* plant.

The present invention also encompasses transcription factor variants. A preferred transcription factor variant is one having at least 40% amino acid sequence identity, a more preferred transcription factor variant is one having at least 50% amino acid sequence identity and a most preferred transcription factor variant is one having at least 65% amino acid sequence identity to the transcription factor amino acid sequence SEQ ID NOs:2 to 2N, where $N = 2-561$, and which contains at least one functional or structural characteristic of the transcription factor amino acid sequence. Sequences having lesser degrees of identity but comparable biological activity are considered to be equivalents.

In another aspect, the invention is a transgenic plant comprising one or more of the above-described isolated or recombinant polynucleotides. In yet another aspect, the invention is a plant with altered expression levels of a polynucleotide described above or a plant with altered expression or activity levels of an above-described polypeptide. Further, the invention is a plant lacking a nucleotide sequence encoding a polypeptide described above or substantially lacking a polypeptide described above. The plant may be any plant, including, but not limited to, *Arabidopsis*, mustard, soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower, alfalfa, sugarcane, turf, banana, blackberry, blueberry, strawberry,

raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, pumpkin, spinach, squash, sweet corn, tobacco, tomato, watermelon, rosaceous fruits, vegetable brassicas, and mint or other labiates. In yet another aspect, the inventions is an isolated plant material of a plant, including, but not limited to, plant tissue, fruit, seed, plant cell, embryo, protoplast, pollen, and the like. In yet another aspect, the invention is a transgenic plant tissue culture of regenerable cells, including, but not limited to, embryos, meristematic cells, microspores, protoplast, pollen, and the like.

In yet another aspect the invention is a transgenic plant comprising one or more of the above described polynucleotides wherein the encoded polypeptide is expressed and regulates transcription of a gene.

In a further aspect the invention provides a method of using the polynucleotide composition to breed a progeny plant from a transgenic plant including crossing plants, producing seeds from transgenic plants, and methods of breeding using transgenic plants, the method comprising transforming a plant with the polynucleotide composition to create a transgenic plant, crossing the transgenic plant with another plant, selecting seed, and growing the progeny plant from the seed.

In a further aspect, the invention provides a progeny plant derived from a parental plant wherein said progeny plant exhibits at least three fold greater messenger RNA levels than said parental plant, wherein the messenger RNA encodes a DNA-binding protein which is capable of binding to a DNA regulatory sequence and inducing expression of a plant trait gene, wherein the progeny plant is characterized by a change in the plant trait compared to said parental plant. In yet a further aspect, the progeny plant exhibits at least ten fold greater messenger RNA levels compared to said parental plant. In yet a further aspect, the progeny plant exhibits at least fifty fold greater messenger RNA levels compared to said parental plant.

In a further aspect, the invention relates to a cloning or expression vector comprising the isolated or recombinant polynucleotide described above or cells comprising the cloning or expression vector.

In yet a further aspect, the invention relates to a composition produced by incubating a polynucleotide of the invention with a nuclease, a restriction enzyme, a polymerase; a polymerase and a primer; a cloning vector, or with a cell.

Furthermore, the invention relates to a method for producing a plant having a modified trait. The method comprises altering the expression of an isolated or recombinant polynucleotide of the invention or altering the expression or activity of a polypeptide of the invention in a plant to produce a modified plant, and selecting the modified plant for a modified trait. In one aspect, the plant is a monocot plant. In another aspect, the plant is a dicot plant. In another aspect the recombinant polynucleotide is from a dicot plant and the plant is a monocot plant. In yet another aspect the recombinant polynucleotide is from a monocot plant and the plant is a dicot plant. In yet another aspect the recombinant polynucleotide is from a monocot plant and the plant is a monocot plant. In yet another aspect the recombinant polynucleotide is from a dicot plant and the plant is a dicot plant.

In another aspect, the invention is a transgenic plant comprising an isolated or recombinant polynucleotide encoding a polypeptide wherein the polypeptide is selected from the group consisting of SEQ ID NOs: 2 - 2N, where $N = 2-561$. In yet another aspect, the invention is a plant with altered expression levels of a polypeptide described above or a plant with altered expression or activity levels of an above-described polypeptide. Further, the invention is a plant lacking a polynucleotide sequence encoding a polypeptide described above or substantially lacking a polypeptide described above. The plant may be any plant, including, but not limited to, Arabidopsis, mustard, soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower, alfalfa, sugarcane, turf, banana, blackberry, blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, pumpkin, spinach, squash, sweet corn, tobacco, tomato, watermelon, rosaceous fruits, vegetable brassicas, and mint or other labiates. In yet another aspect, the inventions is an isolated plant material of a plant, including, but not limited to, plant tissue, fruit, seed, plant cell, embryo, protoplast, pollen, and the like. In yet another aspect, the

invention is a transgenic plant tissue culture of regenerable cells, including, but not limited to, embryos, meristematic cells, microspores, protoplast, pollen, and the like.

In another aspect, the invention relates to a method of identifying a factor that is modulated by or interacts with a polypeptide encoded by a polynucleotide of the invention. The method comprises expressing a polypeptide encoded by the polynucleotide in a plant; and identifying at least one factor that is modulated by or interacts with the polypeptide. In one embodiment the method for identifying modulating or interacting factors is by detecting binding by the polypeptide to a promoter sequence, or by detecting interactions between an additional protein and the polypeptide in a yeast two hybrid system, or by detecting expression of a factor by hybridization to a microarray, subtractive hybridization, or differential display.

In yet another aspect, the invention is a method of identifying a molecule that modulates activity or expression of a polynucleotide or polypeptide of interest. The method comprises placing the molecule in contact with a plant comprising the polynucleotide or polypeptide encoded by the polynucleotide of the invention and monitoring one or more of the expression level of the polynucleotide in the plant, the expression level of the polypeptide in the plant, and modulation of an activity of the polypeptide in the plant.

In yet another aspect, the invention relates to an integrated system, computer or computer readable medium comprising one or more character strings corresponding to a polynucleotide of the invention, or to a polypeptide encoded by the polynucleotide. The integrated system, computer or computer readable medium may comprise a link between one or more sequence strings to a modified plant trait.

In yet another aspect, the invention is a method for identifying a sequence similar or homologous to one or more polynucleotides of the invention, or one or more polypeptides encoded by the polynucleotides. The method comprises providing a sequence database, and querying the sequence database with one or more target sequences corresponding to the one or more polynucleotides or to the one or more polypeptides to identify one or more sequence members of the database that display sequence similarity or homology to one or more of the one or more target sequences.

The method may further comprise of linking the one or more of the polynucleotides of the invention, or encoded polypeptides, to a modified plant phenotype.

BRIEF DESCRIPTION OF THE SEQUENCE LISTING, TABLES, AND FIGURE

The Sequence Listing provides exemplary polynucleotide and polypeptide sequences of the invention. The traits associated with the use of the sequences are included in the Examples.

Diskette1 is a read-only memory computer-readable diskette and contains a copy of the Sequence Listing in ASCII text format. The Sequence Listing is named "SEQLIST514442002041" and is 929 kilobytes in size. The copy of the Sequence Listing on the diskette is hereby incorporated by reference in its entirety.

Table 4 shows the polynucleotides and polypeptides identified by SEQ ID NO; Mendel Gene ID No.; conserved domain of the polypeptide; and if the polynucleotide was tested in a transgenic assay. The first column shows the polynucleotide SEQ ID NO; the second column shows the Mendel Gene ID No., GID; the third column shows the trait(s) resulting from the knock out or overexpression of the polynucleotide in the transgenic plant; the fourth column shows the category of the trait; the fifth column shows the transcription factor family to which the polynucleotide belongs; the sixth column ("Comment"), includes specific effects and utilities conferred by the polynucleotide of the first column; the seventh column shows the SEQ ID NO of the polypeptide encoded by the polynucleotide; and the eighth column shows the amino acid residue positions of the conserved domain in amino acid (AA) co-ordinates.

Table 5 lists a summary of orthologous and homologous sequences identified using BLAST (tblastx program). The first column shows the polynucleotide sequence identifier (SEQ ID NO), the second column shows the corresponding cDNA identifier (Gene ID), the third column shows the orthologous or homologous polynucleotide GenBank Accession Number (Test Sequence ID), the fourth column shows the

calculated probability value that the sequence identity is due to chance (Smallest Sum Probability), the fifth column shows the plant species from which the test sequence was isolated (Test Sequence Species), and the sixth column shows the orthologous or homologous test sequence GenBank annotation (Test Sequence GenBank Annotation).

Figure 1 shows a phylogenic tree of related plant families adapted from Daly et al. (2001 *Plant Physiology* 127:1328-1333).

Detailed Description of Exemplary Embodiments

In an important aspect, the present invention relates to polynucleotides and polypeptides, e.g. for modifying phenotypes of plants. Throughout this disclosure, various information sources are referred to and/or are specifically incorporated. The information sources include scientific journal articles, patent documents, textbooks, and World Wide Web browser-inactive page addresses, for example. While the reference to these information sources clearly indicates that they can be used by one of skill in the art, applicants specifically incorporate each and every one of the information sources cited herein, in their entirety, whether or not a specific mention of "incorporation by reference" is noted. The contents and teachings of each and every one of the information sources can be relied on and used to make and use embodiments of the invention.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a plant" includes a plurality of such plants, and a reference to "a stress" is a reference to one or more stresses and equivalents thereof known to those skilled in the art, and so forth.

The polynucleotide sequences of the invention encode polypeptides that are members of well-known transcription factor families, including plant transcription factor families, as disclosed in Table 4. Generally, the transcription factors encoded by the present sequences are involved in cell differentiation and proliferation and the regulation of growth. Accordingly, one skilled in the art would recognize that by expressing the present sequences in a plant, one may change the expression of

autologous genes or induce the expression of introduced genes. By affecting the expression of similar autologous sequences in a plant that have the biological activity of the present sequences, or by introducing the present sequences into a plant, one may alter a plant's phenotype to one with improved traits. The sequences of the invention may also be used to transform a plant and introduce desirable traits not found in the wild-type cultivar or strain. Plants may then be selected for those that produce the most desirable degree of over- or underexpression of target genes of interest and coincident trait improvement.

The sequences of the present invention may be from any species, particularly plant species, in a naturally occurring form or from any source whether natural, synthetic, semi-synthetic or recombinant. The sequences of the invention may also include fragments of the present amino acid sequences. In this context, a "fragment" refers to a fragment of a polypeptide sequence which is at least 5 to about 15 amino acids in length, most preferably at least 14 amino acids, and which retain some biological activity of a transcription factor. Where "amino acid sequence" is recited to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

As one of ordinary skill in the art recognizes, transcription factors can be identified by the presence of a region or domain of structural similarity or identity to a specific consensus sequence or the presence of a specific consensus DNA-binding site or DNA-binding site motif (see, for example, Riechmann et al., (2000) *Science* 290: 2105-2110). The plant transcription factors may belong to one of the following transcription factor families: the AP2 (APETALA2) domain transcription factor family (Riechmann and Meyerowitz (1998) *Biol. Chem.* 379:633-646); the MYB transcription factor family (Martin and Paz-Ares, (1997) *Trends Genet.* 13:67-73); the MADS domain transcription factor family (Riechmann and Meyerowitz (1997) *Biol. Chem.* 378:1079-1101); the WRKY protein family (Ishiguro and Nakamura (1994) *Mol. Gen. Genet.* 244:563-571); the ankyrin-repeat protein family (Zhang et al. (1992) *Plant Cell* 4:1575-1588); the zinc finger protein (Z) family (Klug and Schwabe (1995) *FASEB J.* 9: 597-604); the homeobox (HB) protein family (Buerklin in *Guidebook to the Homeobox Genes*, Duboule (ed.) (1994) Oxford University Press);

the CAAT-element binding proteins (Forsburg and Guarente (1989) *Genes Dev.* 3:1166-1178); the squamosa promoter binding proteins (SPB) (Klein et al. (1996) *Mol. Gen. Genet.* 1996 250:7-16); the NAM protein family (Souer et al. (1996) *Cell* 85:159-170); the IAA/AUX proteins (Rouse et al. (1998) *Science* 279:1371-1373); the HLH/MYC protein family (Littlewood et al. (1994) *Prot. Profile* 1:639-709); the DNA-binding protein (DBP) family (Tucker et al. (1994) *EMBO J.* 13:2994-3002); the bZIP family of transcription factors (Foster et al. (1994) *FASEB J.* 8:192-200); the Box P-binding protein (the BPF-1) family (da Costa e Silva et al. (1993) *Plant J.* 4:125-135); the high mobility group (HMG) family (Bustin and Reeves (1996) *Prog. Nucl. Acids Res. Mol. Biol.* 54:35-100); the scarecrow (SCR) family (Di Laurenzio et al. (1996) *Cell* 86:423-433); the GF14 family (Wu et al. (1997) *Plant Physiol.* 114:1421-1431); the polycomb (PCOMB) family (Kennison (1995) *Annu. Rev. Genet.* 29:289-303); the teosinte branched (TEO) family (Luo et al. (1996) *Nature* 383:794-799); the ABI3 family (Giraudat et al. (1992) *Plant Cell* 4:1251-1261); the triple helix (TH) family (Dehesh et al. (1990) *Science* 250:1397-1399); the EIL family (Chao et al. (1997) *Cell* 89:1133-44); the AT-HOOK family (Reeves and Nissen (1990) *J. Biol. Chem.* 265:8573-8582); the SIFA family (Zhou et al. (1995) *Nucleic Acids Res.* 23:1165-1169); the bZIPT2 family (Lu and Ferl (1995) *Plant Physiol.* 109:723); the YABBY family (Bowman et al. (1999) *Development* 126:2387-96); the PAZ family (Bohmert et al. (1998) *EMBO J.* 17:170-80); a family of miscellaneous (MISC) transcription factors including the DPBF family (Kim et al. (1997) *Plant J.* 11:1237-1251) and the SPF1 family (Ishiguro and Nakamura (1994) *Mol. Gen. Genet.* 244:563-571); the golden (GLD) family (Hall et al. (1998) *Plant Cell* 10:925-936), the TUBBY family (Boggin et al, (1999) *Science* 286:2119-2125), the heat shock family (Wu C (1995) *Annu Rev Cell Dev Biol* 11:441-469), the ENBP family (Christiansen et al (1996) *Plant Mol Biol* 32:809-821), the RING-zinc family (Jensen et al. (1998) *FEBS letters* 436:283-287), the PDBP family (Janik et al *Virology.* (1989) 168:320-329), the PCF family (Cubas P, et al. *Plant J.* (1999) 18:215-22), the SRS (SHI-related) family (Fridborg et al *Plant Cell* (1999) 11:1019-1032), the CPP (cysteine-rich polycomb-like) family (Cvitanich et al *Proc. Natl. Acad. Sci. U S A.* (2000) 97:8163-8168), the ARF (auxin response factor) family (Ulmasov, et al. (1999) *Proc. Natl. Acad. Sci. USA* 96: 5844-5849), the SWI/SNF family (Collingwood et al *J. Mol. End.* 23:255-275), the ACBF family (Seguin et al (1997) *Plant Mol Biol.* 35:281-291), PCGL (CG-1 like) family (da Costa e Silva et al.

(1994) *Plant Mol Biol.* 25:921-924) the ARID family (Vazquez et al. (1999) *Development.* 126: 733-42), the Jumonji family, Balciunas et al (2000, *Trends Biochem Sci.* 25: 274-276), the bZIP-NIN family (Schauser et al (1999) *Nature* 402: 191-195), the E2F family Kaelin et al (1992) *Cell* 70: 351-364) and the GRF-like family (Knaap et al (2000) *Plant Physiol.* 122: 695-704). As indicated by any part of the list above and as known in the art, transcription factors have been sometimes categorized by class, family, and sub-family according to their structural content and consensus DNA-binding site motif, for example. Many of the classes and many of the families and sub-families are listed here. However, the inclusion of one sub-family and not another, or the inclusion of one family and not another, does not mean that the invention does not encompass polynucleotides or polypeptides of a certain family or sub-family. The list provided here is merely an example of the types of transcription factors and the knowledge available concerning the consensus sequences and consensus DNA-binding site motifs that help define them as known to those of skill in the art (each of the references noted above are specifically incorporated herein by reference). A transcription factor may include, but is not limited to, any polypeptide that can activate or repress transcription of a single gene or a number of genes. This polypeptide group includes, but is not limited to, DNA-binding proteins, DNA-binding protein binding proteins, protein kinases, protein phosphatases, GTP-binding proteins, and receptors, and the like.

In addition to methods for modifying a plant phenotype by employing one or more polynucleotides and polypeptides of the invention described herein, the polynucleotides and polypeptides of the invention have a variety of additional uses. These uses include their use in the recombinant production (i.e., expression) of proteins; as regulators of plant gene expression, as diagnostic probes for the presence of complementary or partially complementary nucleic acids (including for detection of natural coding nucleic acids); as substrates for further reactions, e.g., mutation reactions, PCR reactions, or the like; as substrates for cloning e.g., including digestion or ligation reactions; and for identifying exogenous or endogenous modulators of the transcription factors. A "polynucleotide" is a nucleic acid sequence comprising a plurality of polymerized nucleotides, e.g., at least about 15 consecutive polymerized nucleotides, optionally at least about 30 consecutive nucleotides, at least about 50 consecutive nucleotides. In many instances, a polynucleotide comprises a nucleotide

sequence encoding a polypeptide (or protein) or a domain or fragment thereof. Additionally, the polynucleotide may comprise a promoter, an intron, an enhancer region, a polyadenylation site, a translation initiation site, 5' or 3' untranslated regions, a reporter gene, a selectable marker, or the like. The polynucleotide can be single stranded or double stranded DNA or RNA. The polynucleotide optionally comprises modified bases or a modified backbone. The polynucleotide can be, e.g., genomic DNA or RNA, a transcript (such as an mRNA), a cDNA, a PCR product, a cloned DNA, a synthetic DNA or RNA, or the like. The polynucleotide can comprise a sequence in either sense or antisense orientations.

A "recombinant polynucleotide" is a polynucleotide that is not in its native state, e.g., the polynucleotide comprises a nucleotide sequence not found in nature, or the polynucleotide is in a context other than that in which it is naturally found, e.g., separated from nucleotide sequences with which it typically is in proximity in nature, or adjacent (or contiguous with) nucleotide sequences with which it typically is not in proximity. For example, the sequence at issue can be cloned into a vector, or otherwise recombined with one or more additional nucleic acid.

An "isolated polynucleotide" is a polynucleotide whether naturally occurring or recombinant, that is present outside the cell in which it is typically found in nature, whether purified or not. Optionally, an isolated polynucleotide is subject to one or more enrichment or purification procedures, e.g., cell lysis, extraction, centrifugation, precipitation, or the like.

A "polypeptide" is an amino acid sequence comprising a plurality of consecutive polymerized amino acid residues e.g., at least about 15 consecutive polymerized amino acid residues, optionally at least about 30 consecutive polymerized amino acid residues, at least about 50 consecutive polymerized amino acid residues. In many instances, a polypeptide comprises a polymerized amino acid residue sequence that is a transcription factor or a domain or portion or fragment thereof. Additionally, the polypeptide may comprise a localization domain, 2) an activation domain, 3) a repression domain, 4) an oligomerization domain or 5) a DNA-binding domain, or the like. The polypeptide optionally comprises modified

amino acid residues, naturally occurring amino acid residues not encoded by a codon, non-naturally occurring amino acid residues.

A "recombinant polypeptide" is a polypeptide produced by translation of a recombinant polynucleotide. A "synthetic polypeptide" is a polypeptide created by consecutive polymerization of isolated amino acid residues using methods well known in the art. An "isolated polypeptide," whether a naturally occurring or a recombinant polypeptide, is more enriched in (or out of) a cell than the polypeptide in its natural state in a wild type cell, e.g., more than about 5% enriched, more than about 10% enriched, or more than about 20%, or more than about 50%, or more, enriched, i.e., alternatively denoted: 105%, 110%, 120%, 150% or more, enriched relative to wild type standardized at 100%. Such an enrichment is not the result of a natural response of a wild type plant. Alternatively, or additionally, the isolated polypeptide is separated from other cellular components with which it is typically associated, e.g., by any of the various protein purification methods herein.

"Identity" or "similarity" refers to sequence similarity between two polynucleotide sequences or between two polypeptide sequences, with identity being a more strict comparison. The phrases "percent identity" and "% identity" refer to the percentage of sequence similarity found in a comparison of two or more polynucleotide sequences or two or more polypeptide sequences. Identity or similarity can be determined by comparing a position in each sequence that may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same nucleotide base or amino acid, then the molecules are identical at that position. A degree of similarity or identity between polynucleotide sequences is a function of the number of identical or matching nucleotides at positions shared by the polynucleotide sequences. A degree of identity of polypeptide sequences is a function of the number of identical amino acids at positions shared by the polypeptide sequences. A degree of homology or similarity of polypeptide sequences is a function of the number of amino acids, i.e., structurally related, at positions shared by the polypeptide sequences.

"Altered" nucleic acid sequences encoding polypeptide include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting

in a polynucleotide encoding a polypeptide with at least one functional characteristic of the polypeptide. Included within this definition are polymorphisms that may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding polypeptide, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding polypeptide. The encoded polypeptide protein may also be "altered", and may contain deletions, insertions, or substitutions of amino acid residues that produce a silent change and result in a functionally equivalent polypeptide. Deliberate amino acid substitutions may be made on the basis of similarity in residue side chain chemistry, including, but not limited to, polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological activity of polypeptide is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine. Alignments between different polypeptide sequences may be used to calculate "percentage sequence similarity".

The term "plant" includes whole plants, shoot vegetative organs/structures (e.g., leaves, stems and tubers), roots, flowers and floral organs/structures (e.g., bracts, sepals, petals, stamens, carpels, anthers and ovules), seed (including embryo, endosperm, and seed coat) and fruit (the mature ovary), plant tissue (e.g., vascular tissue, ground tissue, and the like) and cells (e.g., guard cells, egg cells, and the like), and progeny of same. The class of plants that can be used in the method of the invention is generally as broad as the class of higher and lower plants amenable to transformation techniques, including angiosperms (monocotyledonous and dicotyledonous plants), gymnosperms, ferns, horsetails, psilophytes, lycophytes, bryophytes, and multicellular algae. (See for example, Figure 1, adapted from Daly et al. 2001 *Plant Physiology* 127:1328-1333; and see also Tudge, C., The Variety of Life, Oxford University Press, New York, 2000, pp. 547-606.)

A "transgenic plant" refers to a plant that contains genetic material not found in a wild type plant of the same species, variety or cultivar. The genetic material may

include a transgene, an insertional mutagenesis event (such as by transposon or T-DNA insertional mutagenesis), an activation tagging sequence, a mutated sequence, a homologous recombination event or a sequence modified by chimeraplasty. Typically, the foreign genetic material has been introduced into the plant by human manipulation, but any method can be used as one of skill in the art recognizes.

A transgenic plant may contain an expression vector or cassette. The expression cassette typically comprises a polypeptide-encoding sequence operably linked (i.e., under regulatory control of) to appropriate inducible or constitutive regulatory sequences that allow for the expression of polypeptide. The expression cassette can be introduced into a plant by transformation or by breeding after transformation of a parent plant. A plant refers to a whole plant as well as to a plant part, such as seed, fruit, leaf, or root, plant tissue, plant cells or any other plant material, e.g., a plant explant, as well as to progeny thereof, and to *in vitro* systems that mimic biochemical or cellular components or processes in a cell.

"Ectopic expression or altered expression" in reference to a polynucleotide indicates that the pattern of expression in, e.g., a transgenic plant or plant tissue, is different from the expression pattern in a wild type plant or a reference plant of the same species. The pattern of expression may also be compared with a reference expression pattern in a wild type plant of the same species. For example, the polynucleotide or polypeptide is expressed in a cell or tissue type other than a cell or tissue type in which the sequence is expressed in the wild type plant, or by expression at a time other than at the time the sequence is expressed in the wild type plant, or by a response to different inducible agents, such as hormones or environmental signals, or at different expression levels (either higher or lower) compared with those found in a wild type plant. The term also refers to altered expression patterns that are produced by lowering the levels of expression to below the detection level or completely abolishing expression. The resulting expression pattern can be transient or stable, constitutive or inducible. In reference to a polypeptide, the term "ectopic expression or altered expression" further may relate to altered activity levels resulting from the interactions of the polypeptides with exogenous or endogenous modulators or from interactions with factors or as a result of the chemical modification of the polypeptides.

A "fragment" or "domain," with respect to a polypeptide, refers to a subsequence of the polypeptide. In some cases, the fragment or domain, is a subsequence of the polypeptide which performs at least one biological function of the intact polypeptide in substantially the same manner, or to a similar extent, as does the intact polypeptide. For example, a polypeptide fragment can comprise a recognizable structural motif or functional domain such as a DNA-binding site or domain that binds to a DNA promoter region, an activation domain, or a domain for protein-protein interactions. Fragments can vary in size from as few as 6 amino acids to the full length of the intact polypeptide, but are preferably at least about 30 amino acids in length and more preferably at least about 60 amino acids in length. In reference to a polynucleotide sequence, "a fragment" refers to any subsequence of a polynucleotide, typically, of at least about 15 consecutive nucleotides, preferably at least about 30 nucleotides, more preferably at least about 50 nucleotides, of any of the sequences provided herein.

The invention also encompasses production of DNA sequences that encode transcription factors and transcription factor derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding transcription factors or any fragment thereof.

A "conserved domain", with respect to a polypeptide, refers to a domain within a transcription factor family which exhibits a higher degree of sequence homology, such as at least 65% sequence identity including conservative substitutions, and preferably at least 80% sequence identity, and more preferably at least 85%, or at least about 86%, or at least about 87%, or at least about 88%, or at least about 90%, or at least about 95%, or at least about 98% amino acid residue sequence identity of a polypeptide of consecutive amino acid residues. A fragment or domain can be referred to as outside a consensus sequence or outside a consensus DNA-binding site that is known to exist or that exists for a particular transcription factor class, family, or sub-family. In this case, the fragment or domain will not include the exact amino acids of a consensus sequence or consensus DNA-binding

site of a transcription factor class, family or sub-family, or the exact amino acids of a particular transcription factor consensus sequence or consensus DNA-binding site. Furthermore, a particular fragment, region, or domain of a polypeptide, or a polynucleotide encoding a polypeptide, can be "outside a conserved domain" if all the amino acids of the fragment, region, or domain fall outside of a defined conserved domain(s) for a polypeptide or protein. The conserved domains for each of polypeptides of SEQ ID NOs:2 - 2N, where N = 2-561, are listed in Table 4 as described in Example VII. Also, many of the polypeptides of Table 4 have conserved domains specifically indicated by start and stop sites. A comparison of the regions of the polypeptides in SEQ ID NOs:2 - 2N, where N = 2-561, or of those in Table 4, allows one of skill in the art to identify conserved domain(s) for any of the polypeptides listed or referred to in this disclosure, including those in Table 4.

A "trait" refers to a physiological, morphological, biochemical, or physical characteristic of a plant or particular plant material or cell. In some instances, this characteristic is visible to the human eye, such as seed or plant size, or can be measured by biochemical techniques, such as detecting the protein, starch, or oil content of seed or leaves, or by observation of a metabolic or physiological process, e.g. by measuring uptake of carbon dioxide, or by the observation of the expression level of a gene or genes, e.g., by employing Northern analysis, RT-PCR, microarray gene expression assays, or reporter gene expression systems, or by agricultural observations such as stress tolerance, yield, or pathogen tolerance. Any technique can be used to measure the amount of, comparative level of, or difference in any selected chemical compound or macromolecule in the transgenic plants, however.

"Trait modification" refers to a detectable difference in a characteristic in a plant ectopically expressing a polynucleotide or polypeptide of the present invention relative to a plant not doing so, such as a wild type plant. In some cases, the trait modification can be evaluated quantitatively. For example, the trait modification can entail at least about a 2% increase or decrease in an observed trait (difference); at least a 5% difference, at least about a 10% difference, at least about a 20% difference, at least about a 30%, at least about a 50%, at least about a 70%, or at least about a 100%, or an even greater difference compared with a wild type plant. It is known that there can be a natural variation in the modified trait. Therefore, the trait modification

observed entails a change of the normal distribution of the trait in the plants compared with the distribution observed in wild type plant.

I. Traits Which May Be Modified

Trait modifications of particular interest include those to seed (such as embryo or endosperm), fruit, root, flower, leaf, stem, shoot, seedling or the like, including: enhanced tolerance to environmental conditions including freezing, chilling, heat, drought, water saturation, radiation and ozone; improved tolerance to microbial, fungal or viral diseases; improved tolerance to pest infestations, including nematodes, mollicutes, parasitic higher plants or the like; decreased herbicide sensitivity; improved tolerance of heavy metals or enhanced ability to take up heavy metals; improved growth under poor photoconditions (e.g., low light and/or short day length), or changes in expression levels of genes of interest. Other phenotype that can be modified relate to the production of plant metabolites, such as variations in the production of taxol, tocopherol, tocotrienol, sterols, phytosterols, vitamins, wax monomers, anti-oxidants, amino acids, lignins, cellulose, tannins, prenolipids (such as chlorophylls and carotenoids), glucosinolates, and terpenoids, enhanced or compositionally altered protein or oil production (especially in seeds), or modified sugar (insoluble or soluble) and/or starch composition. Physical plant characteristics that can be modified include cell development (such as the number of trichomes), fruit and seed size and number, yields of plant parts such as stems, leaves, inflorescences, and roots, the stability of the seeds during storage, characteristics of the seed pod (e.g., susceptibility to shattering), root hair length and quantity, internode distances, or the quality of seed coat. Plant growth characteristics that can be modified include growth rate, germination rate of seeds, vigor of plants and seedlings, leaf and flower senescence, male sterility, apomixis, flowering time, flower abscission, rate of nitrogen uptake, osmotic sensitivity to soluble sugar concentrations, biomass or transpiration characteristics, as well as plant architecture characteristics such as apical dominance, branching patterns, number of organs, organ identity, organ shape or size.

II. Transcription Factors Modify Expression Of Endogenous Genes

Expression of genes which encode transcription factors that modify expression of endogenous genes, polynucleotides, and proteins are well known in the art. In addition, transgenic plants comprising isolated polynucleotides encoding transcription

factors may also modify expression of endogenous genes, polynucleotides, and proteins. Examples include Peng et al. (1997, *Genes and Development* 11:3194-3205) and Peng et al. (1999, *Nature*, 400:256-261). In addition, many others have demonstrated that an *Arabidopsis* transcription factor expressed in an exogenous plant species elicits the same or very similar phenotypic response. See, for example, Fu et al. (2001, *Plant Cell* 13:1791-1802); Nandi et al. (2000, *Curr. Biol.* 10:215-218); Coupland (1995, *Nature* 377:482-483); and Weigel and Nilsson (1995, *Nature* 377:482-500).

In another example, Mandel et al. (1992, *Cell* 71:133-143) and Suzuki et al. (2001, *Plant J.* 28:409-418) teach that a transcription factor expressed in another plant species elicits the same or very similar phenotypic response of the endogenous sequence, as often predicted in earlier studies of *Arabidopsis* transcription factors in *Arabidopsis* (see Mandel et al., 1992, *supra*; Suzuki et al., 2001, *supra*).

Other examples include Müller et al. (2001, *Plant J.* 28:169-179); Kim et al. (2001, *Plant J.* 25:247-259); Kyoizuka and Shimamoto (2002, *Plant Cell Physiol.* 43:130-135); Boss and Thomas (2002, *Nature*, 416:847-850); He et al. (2000, *Transgenic Res.*, 9:223-227); and Robson et al. (2001, *Plant J.* 28:619-631).

In yet another example, Gilmour et al. (1998, *Plant J.* 16:433-442) teach an *Arabidopsis* AP2 transcription factor, CBF1, which, when overexpressed in transgenic plants, increases plant freezing tolerance. Jaglo et al (2001, *Plant Physiol.* 127:910-917) further identified sequences in *Brassica napus* which encode CBF-like genes and that transcripts for these genes accumulated rapidly in response to low temperature. Transcripts encoding CBF-like proteins were also found to accumulate rapidly in response to low temperature in wheat, as well as in tomato. An alignment of the CBF proteins from *Arabidopsis*, *B. napus*, wheat, rye, and tomato revealed the presence of conserved amino acid sequences, PKK/RPAGR_xKFxETRHP and DSAWR, that bracket the AP2/EREBP DNA binding domains of the proteins and distinguish them from other members of the AP2/EREBP protein family. (See Jaglo et al., *supra*.)

III. Polypeptides and Polynucleotides of the Invention

The present invention provides, among other things, transcription factors (TFs), and transcription factor homologue polypeptides, and isolated or recombinant polynucleotides encoding the polypeptides, or novel variant polypeptides or polynucleotides encoding novel variants of transcription factors derived from the specific sequences provided here. These polypeptides and polynucleotides may be employed to modify a plant's characteristic.

Exemplary polynucleotides encoding the polypeptides of the invention were identified in the *Arabidopsis thaliana* GenBank database using publicly available sequence analysis programs and parameters. Sequences initially identified were then further characterized to identify sequences comprising specified sequence strings corresponding to sequence motifs present in families of known transcription factors. In addition, further exemplary polynucleotides encoding the polypeptides of the invention were identified in the plant GenBank database using publicly available sequence analysis programs and parameters. Sequences initially identified were then further characterized to identify sequences comprising specified sequence strings corresponding to sequence motifs present in families of known transcription factors. Polynucleotide sequences meeting such criteria were confirmed as transcription factors.

Additional polynucleotides of the invention were identified by screening *Arabidopsis thaliana* and/or other plant cDNA libraries with probes corresponding to known transcription factors under low stringency hybridization conditions. Additional sequences, including full length coding sequences were subsequently recovered by the rapid amplification of cDNA ends (RACE) procedure, using a commercially available kit according to the manufacturer's instructions. Where necessary, multiple rounds of RACE are performed to isolate 5' and 3' ends. The full length cDNA was then recovered by a routine end-to-end polymerase chain reaction (PCR) using primers specific to the isolated 5' and 3' ends. Exemplary sequences are provided in the Sequence Listing.

The polynucleotides of the invention can be or were ectopically expressed in overexpressor or knockout plants and the changes in the characteristic(s) or trait(s) of

the plants observed. Therefore, the polynucleotides and polypeptides can be employed to improve the characteristics of plants.

The polynucleotides of the invention can be or were ectopically expressed in overexpressor plant cells and the changes in the expression levels of a number of genes, polynucleotides, and/or proteins of the plant cells observed. Therefore, the polynucleotides and polypeptides can be employed to change expression levels of a genes, polynucleotides, and/or proteins of plants.

IV. Producing Polypeptides

The polynucleotides of the invention include sequences that encode transcription factors and transcription factor homologue polypeptides and sequences complementary thereto, as well as unique fragments of coding sequence, or sequence complementary thereto. Such polynucleotides can be, e.g., DNA or RNA, e.g., mRNA, cRNA, synthetic RNA, genomic DNA, cDNA synthetic DNA, oligonucleotides, etc. The polynucleotides are either double-stranded or single-stranded, and include either, or both sense (i.e., coding) sequences and antisense (i.e., non-coding, complementary) sequences. The polynucleotides include the coding sequence of a transcription factor, or transcription factor homologue polypeptide, in isolation, in combination with additional coding sequences (e.g., a purification tag, a localization signal, as a fusion-protein, as a pre-protein, or the like), in combination with non-coding sequences (e.g., introns or inteins, regulatory elements such as promoters, enhancers, terminators, and the like), and/or in a vector or host environment in which the polynucleotide encoding a transcription factor or transcription factor homologue polypeptide is an endogenous or exogenous gene.

A variety of methods exist for producing the polynucleotides of the invention. Procedures for identifying and isolating DNA clones are well known to those of skill in the art, and are described in, e.g., Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volume 152 Academic Press, Inc., San Diego, CA ("Berger"); Sambrook et al., Molecular Cloning - A Laboratory Manual (2nd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989 ("Sambrook") and Current Protocols in Molecular Biology, F. M. Ausubel et al., eds.,

Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 2000) ("Ausubel").

Alternatively, polynucleotides of the invention, can be produced by a variety of in vitro amplification methods adapted to the present invention by appropriate selection of specific or degenerate primers. Examples of protocols sufficient to direct persons of skill through in vitro amplification methods, including the polymerase chain reaction (PCR) the ligase chain reaction (LCR), Qbeta-replicase amplification and other RNA polymerase mediated techniques (e.g., NASBA), e.g., for the production of the homologous nucleic acids of the invention are found in Berger (supra), Sambrook (supra), and Ausubel (supra), as well as Mullis et al., (1987) PCR Protocols A Guide to Methods and Applications (Innis et al. eds) Academic Press Inc. San Diego, CA (1990) (Innis). Improved methods for cloning in vitro amplified nucleic acids are described in Wallace et al., U.S. Pat. No. 5,426,039. Improved methods for amplifying large nucleic acids by PCR are summarized in Cheng et al. (1994) Nature 369: 684-685 and the references cited therein, in which PCR amplicons of up to 40kb are generated. One of skill will appreciate that essentially any RNA can be converted into a double stranded DNA suitable for restriction digestion, PCR expansion and sequencing using reverse transcriptase and a polymerase. See, e.g., Ausubel, Sambrook and Berger, *all supra*.

Alternatively, polynucleotides and oligonucleotides of the invention can be assembled from fragments produced by solid-phase synthesis methods. Typically, fragments of up to approximately 100 bases are individually synthesized and then enzymatically or chemically ligated to produce a desired sequence, e.g., a polynucleotide encoding all or part of a transcription factor. For example, chemical synthesis using the phosphoramidite method is described, e.g., by Beaucage et al. (1981) Tetrahedron Letters 22:1859-1869; and Matthes et al. (1984) EMBO J. 3:801-805. According to such methods, oligonucleotides are synthesized, purified, annealed to their complementary strand, ligated and then optionally cloned into suitable vectors. And if so desired, the polynucleotides and polypeptides of the invention can be custom ordered from any of a number of commercial suppliers.

V. Homologous Sequences

Sequences homologous, i.e., that share significant sequence identity or similarity, to those provided in the Sequence Listing, derived from *Arabidopsis thaliana* or from other plants of choice are also an aspect of the invention.

Homologous sequences can be derived from any plant including monocots and dicots and in particular agriculturally important plant species, including but not limited to, crops such as soybean, wheat, corn, potato, cotton, rice, rape, oilseed rape (including canola), sunflower, alfalfa, sugarcane and turf; or fruits and vegetables, such as banana, blackberry, blueberry, strawberry, and raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, pumpkin, spinach, squash, sweet corn, tobacco, tomato, watermelon, rosaceous fruits (such as apple, peach, pear, cherry and plum) and vegetable brassicas (such as broccoli, cabbage, cauliflower, Brussels sprouts, and kohlrabi). Other crops, fruits and vegetables whose phenotype can be changed include barley, rye, millet, sorghum, currant, avocado, citrus fruits such as oranges, lemons, grapefruit and tangerines, artichoke, cherries, nuts such as the walnut and peanut, endive, leek, roots, such as arrowroot, beet, cassava, turnip, radish, yam, and sweet potato, and beans. The homologous sequences may also be derived from woody species, such as pine, poplar and eucalyptus, or mint or other labiates.

Orthologs And Paralogs

Several different methods are known by those of skill in the art for identifying and defining these functionally homologous sequences. Three general methods for defining paralogs and orthologs are described; a paralog or ortholog or homolog may be identified by one or more of the methods described below.

Orthologs and paralogs are evolutionarily related genes that have similar sequence and similar functions. Orthologs are structurally related genes in different species that are derived from a speciation event. Paralogs are structurally related genes within a single species that are derived by a duplication event.

Within a single plant species, gene duplication may cause two copies of a particular gene, giving rise to two or more genes with similar sequence and similar function known as paralogs. A paralog is therefore a similar gene with a similar function within the same species. Paralogs typically cluster together or in the same

clade (a group of similar genes) when a gene family phylogeny is analyzed using programs such as CLUSTAL (Thompson et al. (1994) *Nucleic Acids Res.* 22:4673-4680; Higgins et al. (1996) *Methods Enzymol.* 266 383-402). Groups of similar genes can also be identified with pair-wise BLAST analysis (Feng and Doolittle (1987) *J. Mol. Evol.* 25:351-360). For example, a clade of very similar MADS domain transcription factors from *Arabidopsis* all share a common function in flowering time (Ratcliffe et al. (2001) *Plant Physiol.* 126:122-132), and a group of very similar AP2 domain transcription factors from *Arabidopsis* are involved in tolerance of plants to freezing (Gilmour et al. (1998) *Plant J.* 16:433-442). Analysis of groups of similar genes with similar function that fall within one clade can yield sub-sequences that are particular to the clade. These sub-sequences, known as consensus sequences, can not only be used to define the sequences within each clade, but define the functions of these genes; genes within a clade may contain paralogous or orthologous sequences that share the same function. (See also, for example, Mount, D.W. (2001) Bioinformatics: Sequence and Genome Analysis Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York page 543.)

Speciation, the production of new species from a parental species, can also give rise to two or more genes with similar sequence and similar function. These genes, termed orthologs, often have an identical function within their host plants and are often interchangeable between species without losing function. Because plants have common ancestors, many genes in any plant species will have a corresponding orthologous gene in another plant species. Once a phylogenetic tree for a gene family of one species has been constructed using a program such as CLUSTAL (Thompson et al. (1994) *Nucleic Acids Res.* 22:4673-4680; Higgins et al. (1996) *Methods Enzymol.* 266:383-402), potential orthologous sequences can be placed into the phylogenetic tree and its relationship to genes from the species of interest can be determined. Once the ortholog pair has been identified, the function of the test ortholog can be determined by determining the function of the reference ortholog.

Transcription factors that are homologous to the listed sequences will typically share at least about 30% amino acid sequence identity, or at least about 30% amino acid sequence identity outside of a known consensus sequence or consensus DNA-binding site. More closely related transcription factors can share at least about 50%,

about 60%, about 65%, about 70%, about 75% or about 80% or about 90% or about 95% or about 98% or more sequence identity with the listed sequences, or with the listed sequences but excluding or outside a known consensus sequence or consensus DNA-binding site, or with the listed sequences excluding one or all conserved domain. Factors that are most closely related to the listed sequences share, e.g., at least about 85%, about 90% or about 95% or more % sequence identity to the listed sequences, or to the listed sequences but excluding or outside a known consensus sequence or consensus DNA-binding site or outside one or all conserved domain. At the nucleotide level, the sequences will typically share at least about 40% nucleotide sequence identity, preferably at least about 50%, about 60%, about 70% or about 80% sequence identity, and more preferably about 85%, about 90%, about 95% or about 97% or more sequence identity to one or more of the listed sequences, or to a listed sequence but excluding or outside a known consensus sequence or consensus DNA-binding site, or outside one or all conserved domain. The degeneracy of the genetic code enables major variations in the nucleotide sequence of a polynucleotide while maintaining the amino acid sequence of the encoded protein. Conserved domains within a transcription factor family may exhibit a higher degree of sequence homology, such as at least 65% sequence identity including conservative substitutions, and preferably at least 80% sequence identity, and more preferably at least 85%, or at least about 86%, or at least about 87%, or at least about 88%, or at least about 90%, or at least about 95%, or at least about 98% sequence identity. Transcription factors that are homologous to the listed sequences should share at least 30%, or at least about 60%, or at least about 75%, or at least about 80%, or at least about 90%, or at least about 95% amino acid sequence identity over the entire length of the polypeptide or the homolog. In addition, transcription factors that are homologous to the listed sequences should share at least 30%, or at least about 60%, or at least about 75%, or at least about 80%, or at least about 90%, or at least about 95% amino acid sequence similarity over the entire length of the polypeptide or the homolog.

Percent identity can be determined electronically, e.g., by using the MEGALIGN program (DNASTAR, Inc. Madison, Wis.). The MEGALIGN program can create alignments between two or more sequences according to different methods, e.g., the clustal method. (See, e.g., Higgins, D. G. and P. M. Sharp (1988) Gene

73:237-244.) The clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. Other alignment algorithms or programs may be used, including FASTA, BLAST, or ENTREZ, FASTA and BLAST. These are available as a part of the GCG sequence analysis package (University of Wisconsin, Madison, Wis.), and can be used with or without default settings. ENTREZ is available through the National Center for Biotechnology Information. In one embodiment, the percent identity of two sequences can be determined by the GCG program with a gap weight of 1, e.g., each amino acid gap is weighted as if it were a single amino acid or nucleotide mismatch between the two sequences (see USPN 6,262,333).

Other techniques for alignment are described in *Methods in Enzymology*, vol. 266: *Computer Methods for Macromolecular Sequence Analysis* (1996), ed. Doolittle, Academic Press, Inc., San Diego, Calif., USA. Preferably, an alignment program that permits gaps in the sequence is utilized to align the sequences. The Smith-Waterman is one type of algorithm that permits gaps in sequence alignments. See *Methods Mol. Biol.* 70: 173-187 (1997). Also, the GAP program using the Needleman and Wunsch alignment method can be utilized to align sequences. An alternative search strategy uses MPSRCH software, which runs on a MASPAR computer. MPSRCH uses a Smith-Waterman algorithm to score sequences on a massively parallel computer. This approach improves ability to pick up distantly related matches, and is especially tolerant of small gaps and nucleotide sequence errors. Nucleic acid-encoded amino acid sequences can be used to search both protein and DNA databases.

The percentage similarity between two polypeptide sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no similarity between the two amino acid sequences are not included in determining percentage similarity. Percent identity between polynucleotide sequences can also be counted or calculated by other methods known in the art, e.g., the Jotun Hein method. (See, e.g., Hein, J. (1990) *Methods Enzymol.* 183:626-645.) Identity between sequences can also be determined by other methods

known in the art, e.g., by varying hybridization conditions (see US Patent Application No. 20010010913).

Thus, the invention provides methods for identifying a sequence similar or paralogous or orthologous or homologous to one or more polynucleotides as noted herein, or one or more target polypeptides encoded by the polynucleotides, or otherwise noted herein and may include linking or associating a given plant phenotype or gene function with a sequence. In the methods, a sequence database is provided (locally or across an inter or intra net) and a query is made against the sequence database using the relevant sequences herein and associated plant phenotypes or gene functions.

In addition, one or more polynucleotide sequences or one or more polypeptides encoded by the polynucleotide sequences may be used to search against a BLOCKS (Bairoch et al. (1997) *Nucleic Acids Res.* 25:217-221), PFAM, and other databases which contain previously identified and annotated motifs, sequences and gene functions. Methods that search for primary sequence patterns with secondary structure gap penalties (Smith et al. (1992) *Protein Engineering* 5:35-51) as well as algorithms such as Basic Local Alignment Search Tool (BLAST; Altschul, S. F. (1993) *J. Mol. Evol.* 36:290-300; Altschul et al. (1990) *supra*), BLOCKS (Henikoff, S. and Henikoff, G. J. (1991) *Nucleic Acids Research* 19:6565-6572), Hidden Markov Models (HMM; Eddy, S. R. (1996) *Cur. Opin. Str. Biol.* 6:361-365; Sonnhammer et al. (1997) *Proteins* 28:405-420), and the like, can be used to manipulate and analyze polynucleotide and polypeptide sequences encoded by polynucleotides. These databases, algorithms and other methods are well known in the art and are described in Ausubel et al. (1997; *Short Protocols in Molecular Biology*, John Wiley & Sons, New York N.Y., unit 7.7) and in Meyers, R. A. (1995; *Molecular Biology and Biotechnology*, Wiley VCH, New York N.Y., p 856-853).

Furthermore, methods using manual alignment of sequences similar or homologous to one or more polynucleotide sequences or one or more polypeptides encoded by the polynucleotide sequences may be used to identify regions of similarity and conserved domains. Such manual methods are well-known of those of skill in the art and can include, for example, comparisons of tertiary structure between a

polypeptide sequence encoded by a polynucleotide which comprises a known function with a polypeptide sequence encoded by a polynucleotide sequence which has a function not yet determined. Such examples of tertiary structure may comprise predicted alpha helices, beta-sheets, amphipathic helices, leucine zipper motifs, zinc finger motifs, proline-rich regions, cysteine repeat motifs, and the like.

VI. Identifying Polynucleotides or Nucleic Acids by Hybridization

Polynucleotides homologous to the sequences illustrated in the Sequence Listing and tables can be identified, e.g., by hybridization to each other under stringent or under highly stringent conditions. Single stranded polynucleotides hybridize when they associate based on a variety of well characterized physical-chemical forces, such as hydrogen bonding, solvent exclusion, base stacking and the like. The stringency of a hybridization reflects the degree of sequence identity of the nucleic acids involved, such that the higher the stringency, the more similar are the two polynucleotide strands. Stringency is influenced by a variety of factors, including temperature, salt concentration and composition, organic and non-organic additives, solvents, etc. present in both the hybridization and wash solutions and incubations (and number thereof), as described in more detail in the references cited above. Encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NOs: 860; 802; 240; 274; 558; 24; 1120; 44; 460; 286; 120; 130; 134; 698; 832; 580; 612; 48, and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G. M. and S. L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A. R. (1987) *Methods Enzymol.* 152:507-511.) Estimates of homology are provided by either DNA-DNA or DNA-RNA hybridization under conditions of stringency as is well understood by those skilled in the art (Hames and Higgins, Eds. (1985) *Nucleic Acid Hybridisation*, IRL Press, Oxford, U.K.). Stringency conditions can be adjusted to screen for moderately similar fragments, such as homologous sequences from distantly related organisms, to highly similar fragments, such as genes that duplicate functional enzymes from closely related organisms. Post-hybridization washes determine stringency conditions.

In addition to the nucleotide sequences listed in Tables 4 and 5, full length cDNA, orthologs, paralogs and homologs of the present nucleotide sequences may be

identified and isolated using well known methods. The cDNA libraries orthologs, paralogs and homologs of the present nucleotide sequences may be screened using hybridization methods to determine their utility as hybridization target or amplification probes.

An example of stringent hybridization conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or northern blot is about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Nucleic acid molecules that hybridize under stringent conditions will typically hybridize to a probe based on either the entire cDNA or selected portions, e.g., to a unique subsequence, of the cDNA under wash conditions of 0.2x SSC to 2.0 x SSC, 0.1% SDS at 50-65° C. For example, high stringency is about 0.2 x SSC, 0.1% SDS at 65° C. Ultra-high stringency will be the same conditions except the wash temperature is raised about 3 to about 5° C, and ultra-ultra-high stringency will be the same conditions except the wash temperature is raised about 6 to about 9° C. For identification of less closely related homologues washes can be performed at a lower temperature, e.g., 50° C. In general, stringency is increased by raising the wash temperature and/or decreasing the concentration of SSC, as known in the art.

In another example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and most preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, e.g., formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and most preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30° C, more preferably of at least about 37° C, and most preferably of at least about 42° C. Varying additional parameters, such as hybridization time, the concentration of detergent, e.g., sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the

art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred embodiment, hybridization will occur at 30° C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37° C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 µg/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42° C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50% formamide, and 200 µg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

The washing steps that follow hybridization can also vary in stringency. Wash stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include temperature of at least about 25° C, more preferably of at least about 42° C. Another preferred set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1% SDS at 65° C. The most preferred high stringency washes are of at least about 68° C. For example, in a preferred embodiment, wash steps will occur at 25° C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 42° C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a most preferred embodiment, the wash steps will occur at 68° C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional variations on these conditions will be readily apparent to those skilled in the art (see U.S. Patent Application No. 20010010913).

As another example, stringent conditions can be selected such that an oligonucleotide that is perfectly complementary to the coding oligonucleotide hybridizes to the coding oligonucleotide with at least about a 5-10x higher signal to noise ratio than the ratio for hybridization of the perfectly complementary oligonucleotide to a nucleic acid encoding a transcription factor known as of the filing

date of the application. Conditions can be selected such that a higher signal to noise ratio is observed in the particular assay which is used, e.g., about 15x, 25x, 35x, 50x or more. Accordingly, the subject nucleic acid hybridizes to the unique coding oligonucleotide with at least a 2x higher signal to noise ratio as compared to hybridization of the coding oligonucleotide to a nucleic acid encoding known polypeptide. Again, higher signal to noise ratios can be selected, e.g., about 5x, 10x, 25x, 35x, 50x or more. The particular signal will depend on the label used in the relevant assay, e.g., a fluorescent label, a colorimetric label, a radioactive label, or the like.

Alternatively, transcription factor homolog polypeptides can be obtained by screening an expression library using antibodies specific for one or more transcription factors. With the provision herein of the disclosed transcription factor, and transcription factor homologue nucleic acid sequences, the encoded polypeptide(s) can be expressed and purified in a heterologous expression system (e.g., *E. coli*) and used to raise antibodies (monoclonal or polyclonal) specific for the polypeptide(s) in question. Antibodies can also be raised against synthetic peptides derived from transcription factor, or transcription factor homologue, amino acid sequences. Methods of raising antibodies are well known in the art and are described in Harlow and Lane (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, New York. Such antibodies can then be used to screen an expression library produced from the plant from which it is desired to clone additional transcription factor homologues, using the methods described above. The selected cDNAs can be confirmed by sequencing and enzymatic activity.

VII. Sequence Variations

It will readily be appreciated by those of skill in the art, that any of a variety of polynucleotide sequences are capable of encoding the transcription factors and transcription factor homologue polypeptides of the invention. Due to the degeneracy of the genetic code, many different polynucleotides can encode identical and/or substantially similar polypeptides in addition to those sequences illustrated in the Sequence Listing. Nucleic acids having a sequence that differs from the sequences shown in the Sequence Listing, or complementary sequences, that encode functionally equivalent peptides (i.e., peptides having some degree of equivalent or similar

biological activity) but differ in sequence from the sequence shown in the sequence listing due to degeneracy in the genetic code, are also within the scope of the invention.

Altered polynucleotide sequences encoding polypeptides include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polynucleotide encoding a polypeptide with at least one functional characteristic of the instant polypeptides. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding the instant polypeptides, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding the instant polypeptides.

Allelic variant refers to any of two or more alternative forms of a gene occupying the same chromosomal locus. Allelic variation arises naturally through mutation, and may result in phenotypic polymorphism within populations. Gene mutations can be silent (i.e., no change in the encoded polypeptide) or may encode polypeptides having altered amino acid sequence. The term allelic variant is also used herein to denote a protein encoded by an allelic variant of a gene. Splice variant refers to alternative forms of RNA transcribed from a gene. Splice variation arises naturally through use of alternative splicing sites within a transcribed RNA molecule, or less commonly between separately transcribed RNA molecules, and may result in several mRNAs transcribed from the same gene. Splice variants may encode polypeptides having altered amino acid sequence. The term splice variant is also used herein to denote a protein encoded by a splice variant of an mRNA transcribed from a gene.

Those skilled in the art would recognize that the polypeptide sequence G681, SEQ ID NO: 580, represents a single transcription factor; allelic variation and alternative splicing may be expected to occur. Allelic variants of the polypeptide sequence of SEQ ID NO: 579 can be cloned by probing cDNA or genomic libraries from different individual organisms according to standard procedures. Allelic variants of the DNA sequence shown in SEQ ID NO: 579, including those containing silent mutations and those in which mutations result in amino acid sequence changes, are within the scope of the present invention, as are proteins which are allelic variants

of SEQ ID NO: 580. cDNAs generated from alternatively spliced mRNAs, which retain the properties of the transcription factor are included within the scope of the present invention, as are polypeptides encoded by such cDNAs and mRNAs. Allelic variants and splice variants of these sequences can be cloned by probing cDNA or genomic libraries from different individual organisms or tissues according to standard procedures known in the art (see USPN 6,388,064).

For example, Table 1 illustrates, e.g., that the codons AGC, AGT, TCA, TCC, TCG, and TCT all encode the same amino acid: serine. Accordingly, at each position in the sequence where there is a codon encoding serine, any of the above trinucleotide sequences can be used without altering the encoded polypeptide.

Table 1

Amino acid			Possible Codons							
Alanine	Ala	A	GCA	GCC	GCG	GCU				
Cysteine	Cys	C	TGC	TGT						
Aspartic acid	Asp	D	GAC	GAT						
Glutamic acid	Glu	E	GAA	GAG						
Phenylalanine	Phe	F	TTC	TTT						
Glycine	Gly	G	GGA	GGC	GGG	GGT				
Histidine	His	H	CAC	CAT						
Isoleucine	Ile	I	ATA	ATC	ATT					
Lysine	Lys	K	AAA	AAG						
Leucine	Leu	L	TTA	TTG	CTA	CTC	CTG	CTT		
Methionine	Met	M	ATG							
Asparagine	Asn	N	AAC	AAT						
Proline	Pro	P	CCA	CCC	CCG	CCT				
Glutamine	Gln	Q	CAA	CAG						
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGT		
Serine	Ser	S	AGC	AGT	TCA	TCC	TCG	TCT		
Threonine	Thr	T	ACA	ACC	ACG	ACT				
Valine	Val	V	GTA	GTC	GTG	GTT				
Tryptophan	Trp	W	TGG							
Tyrosine	Tyr	Y	TAC	TAT						

Sequence alterations that do not change the amino acid sequence encoded by the polynucleotide are termed "silent" variations. With the exception of the codons ATG and TGG, encoding methionine and tryptophan, respectively, any of the possible codons for the same amino acid can be substituted by a variety of techniques, e.g., site-directed mutagenesis, available in the art. Accordingly, any and all such variations of a sequence selected from the above table are a feature of the invention.

In addition to silent variations, other conservative variations that alter one, or a few amino acids in the encoded polypeptide, can be made without altering the function of the polypeptide, these conservative variants are, likewise, a feature of the invention.

For example, substitutions, deletions and insertions introduced into the sequences provided in the Sequence Listing are also envisioned by the invention. Such sequence modifications can be engineered into a sequence by site-directed mutagenesis (Wu (ed.) Meth. Enzymol. (1993) vol. 217, Academic Press) or the other methods noted below. Amino acid substitutions are typically of single residues; insertions usually will be on the order of about from 1 to 10 amino acid residues; and deletions will range about from 1 to 30 residues. In preferred embodiments, deletions or insertions are made in adjacent pairs, e.g., a deletion of two residues or insertion of two residues. Substitutions, deletions, insertions or any combination thereof can be combined to arrive at a sequence. The mutations that are made in the polynucleotide encoding the transcription factor should not place the sequence out of reading frame and should not create complementary regions that could produce secondary mRNA structure. Preferably, the polypeptide encoded by the DNA performs the desired function.

Conservative substitutions are those in which at least one residue in the amino acid sequence has been removed and a different residue inserted in its place. Such substitutions generally are made in accordance with the Table 2 when it is desired to maintain the activity of the protein. Table 2 shows amino acids which can be substituted for an amino acid in a protein and which are typically regarded as conservative substitutions.

Table 2

Residue	Conservative Substitutions
Ala	Ser
Arg	Lys
Asn	Gln; His
Asp	Glu
Gln	Asn
Cys	Ser
Glu	Asp
Gly	Pro
His	Asn; Gln
Ile	Leu, Val
Leu	Ile; Val
Lys	Arg; Gln
Met	Leu; Ile
Phe	Met; Leu; Tyr
Ser	Thr; Gly
Thr	Ser; Val
Trp	Tyr
Tyr	Trp; Phe
Val	Ile; Leu

Similar substitutions are those in which at least one residue in the amino acid sequence has been removed and a different residue inserted in its place. Such substitutions generally are made in accordance with the Table 3 when it is desired to maintain the activity of the protein. Table 3 shows amino acids which can be substituted for an amino acid in a protein and which are typically regarded as structural and functional substitutions. For example, a residue in column 1 of Table 3

may be substituted with residue in column 2; in addition, a residue in column 2 of Table 3 may be substituted with the residue of column 1.

Table 3

Residue	Similar Substitutions
Ala	Ser; Thr; Gly; Val; Leu; Ile
Arg	Lys; His; Gly
Asn	Gln; His; Gly; Ser; Thr
Asp	Glu; Ser; Thr
Gln	Asn; Ala
Cys	Ser; Gly
Glu	Asp
Gly	Pro; Arg
His	Asn; Gln; Tyr; Phe; Lys; Arg
Ile	Ala; Leu; Val; Gly; Met
Leu	Ala; Ile; Val; Gly; Met
Lys	Arg; His; Gln; Gly; Pro
Met	Leu; Ile; Phe
Phe	Met; Leu; Tyr; Trp; His; Val; Ala
Ser	Thr; Gly; Asp; Ala; Val; Ile; His
Thr	Ser; Val; Ala; Gly
Trp	Tyr; Phe; His
Tyr	Trp; Phe; His
Val	Ala; Ile; Leu; Gly; Thr; Ser; Glu

Substitutions that are less conservative than those in Table 2 can be selected by picking residues that differ more significantly in their effect on maintaining (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain. The substitutions which in general are

expected to produce the greatest changes in protein properties will be those in which (a) a hydrophilic residue, e.g., seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g., leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g., lysyl, arginyl, or histidyl, is substituted for (or by) an electronegative residue, e.g., glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g., phenylalanine, is substituted for (or by) one not having a side chain, e.g., glycine.

VIII. Further Modifying Sequences of the Invention – Mutation/Forced Evolution

In addition to generating silent or conservative substitutions as noted, above, the present invention optionally includes methods of modifying the sequences of the Sequence Listing. In the methods, nucleic acid or protein modification methods are used to alter the given sequences to produce new sequences and/or to chemically or enzymatically modify given sequences to change the properties of the nucleic acids or proteins.

Thus, in one embodiment, given nucleic acid sequences are modified, e.g., according to standard mutagenesis or artificial evolution methods to produce modified sequences. The modified sequences may be created using purified natural polynucleotides isolated from any organism or may be synthesized from purified compositions and chemicals using chemical means well known to those of skill in the art. For example, Ausubel, *supra*, provides additional details on mutagenesis methods. Artificial forced evolution methods are described, for example, by Stemmer (1994) Nature 370:389-391, Stemmer (1994) Proc. Natl. Acad. Sci. USA 91:10747-10751, and U.S. Patents 5,811,238, 5,837,500, and 6,242,568. Methods for engineering synthetic transcription factors and other polypeptides are described, for example, by Zhang et al. (2000) J. Biol. Chem. 275:33850-33860, Liu et al. (2001) J. Biol. Chem. 276:11323-11334, and Isalan et al. (2001) Nature Biotechnol. 19:656-660. Many other mutation and evolution methods are also available and expected to be within the skill of the practitioner.

Similarly, chemical or enzymatic alteration of expressed nucleic acids and polypeptides can be performed by standard methods. For example, sequence can be modified by addition of lipids, sugars, peptides, organic or inorganic compounds, by the inclusion of modified nucleotides or amino acids, or the like. For example, protein modification techniques are illustrated in Ausubel, *supra*. Further details on chemical and enzymatic modifications can be found herein. These modification methods can be used to modify any given sequence, or to modify any sequence produced by the various mutation and artificial evolution modification methods noted herein.

Accordingly, the invention provides for modification of any given nucleic acid by mutation, evolution, chemical or enzymatic modification, or other available methods, as well as for the products produced by practicing such methods, e.g., using the sequences herein as a starting substrate for the various modification approaches.

For example, optimized coding sequence containing codons preferred by a particular prokaryotic or eukaryotic host can be used e.g., to increase the rate of translation or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life, as compared with transcripts produced using a non-optimized sequence. Translation stop codons can also be modified to reflect host preference. For example, preferred stop codons for *Saccharomyces cerevisiae* and mammals are TAA and TGA, respectively. The preferred stop codon for monocotyledonous plants is TGA, whereas insects and *E. coli* prefer to use TAA as the stop codon.

The polynucleotide sequences of the present invention can also be engineered in order to alter a coding sequence for a variety of reasons, including but not limited to, alterations which modify the sequence to facilitate cloning, processing and/or expression of the gene product. For example, alterations are optionally introduced using techniques which are well known in the art, e.g., site-directed mutagenesis, to insert new restriction sites, to alter glycosylation patterns, to change codon preference, to introduce splice sites, etc.

Furthermore, a fragment or domain derived from any of the polypeptides of the invention can be combined with domains derived from other transcription factors or synthetic domains to modify the biological activity of a transcription factor. For instance, a DNA-binding domain derived from a transcription factor of the invention can be combined with the activation domain of another transcription factor or with a synthetic activation domain. A transcription activation domain assists in initiating transcription from a DNA-binding site. Examples include the transcription activation region of VP16 or GAL4 (Moore et al. (1998) Proc. Natl. Acad. Sci. USA 95: 376-381; and Aoyama et al. (1995) Plant Cell 7:1773-1785), peptides derived from bacterial sequences (Ma and Ptashne (1987) Cell 51: 113-119) and synthetic peptides (Giniger and Ptashne, (1987) Nature 330:670-672).

IX. Expression and Modification of Polypeptides

Typically, polynucleotide sequences of the invention are incorporated into recombinant DNA (or RNA) molecules that direct expression of polypeptides of the invention in appropriate host cells, transgenic plants, in vitro translation systems, or the like. Due to the inherent degeneracy of the genetic code, nucleic acid sequences which encode substantially the same or a functionally equivalent amino acid sequence can be substituted for any listed sequence to provide for cloning and expressing the relevant homologue.

X. Vectors, Promoters, and Expression Systems

The present invention includes recombinant constructs comprising one or more of the nucleic acid sequences herein. The constructs typically comprise a vector, such as a plasmid, a cosmid, a phage, a virus (e.g., a plant virus), a bacterial artificial chromosome (BAC), a yeast artificial chromosome (YAC), or the like, into which a nucleic acid sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available.

General texts that describe molecular biological techniques useful herein, including the use and production of vectors, promoters and many other relevant

topics, include Berger, Sambrook and Ausubel, *supra*. Any of the identified sequences can be incorporated into a cassette or vector, e.g., for expression in plants. A number of expression vectors suitable for stable transformation of plant cells or for the establishment of transgenic plants have been described including those described in Weissbach and Weissbach, (1989) Methods for Plant Molecular Biology, Academic Press, and Gelvin et al., (1990) Plant Molecular Biology Manual, Kluwer Academic Publishers. Specific examples include those derived from a Ti plasmid of *Agrobacterium tumefaciens*, as well as those disclosed by Herrera-Estrella et al. (1983) Nature 303: 209, Bevan (1984) Nucl Acid Res. 12: 8711-8721, Klee (1985) Bio/Technology 3: 637-642, for dicotyledonous plants.

Alternatively, non-Ti vectors can be used to transfer the DNA into monocotyledonous plants and cells by using free DNA delivery techniques. Such methods can involve, for example, the use of liposomes, electroporation, microprojectile bombardment, silicon carbide whiskers, and viruses. By using these methods transgenic plants such as wheat, rice (Christou (1991) Bio/Technology 9: 957-962) and corn (Gordon-Kamm (1990) Plant Cell 2: 603-618) can be produced. An immature embryo can also be a good target tissue for monocots for direct DNA delivery techniques by using the particle gun (Weeks et al. (1993) Plant Physiol 102: 1077-1084; Vasil (1993) Bio/Technology 10: 667-674; Wan and Lemeaux (1994) Plant Physiol 104: 37-48, and for *Agrobacterium*-mediated DNA transfer (Ishida et al. (1996) Nature Biotech 14: 745-750).

Typically, plant transformation vectors include one or more cloned plant coding sequence (genomic or cDNA) under the transcriptional control of 5' and 3' regulatory sequences and a dominant selectable marker. Such plant transformation vectors typically also contain a promoter (e.g., a regulatory region controlling inducible or constitutive, environmentally-or developmentally-regulated, or cell- or tissue-specific expression), a transcription initiation start site, an RNA processing signal (such as intron splice sites), a transcription termination site, and/or a polyadenylation signal.

Examples of constitutive plant promoters which can be useful for expressing the TF sequence include: the cauliflower mosaic virus (CaMV) 35S promoter, which

confers constitutive, high-level expression in most plant tissues (*see, e.g.,* Odell et al. (1985) Nature 313:810-812); the nopaline synthase promoter (An et al. (1988) Plant Physiol 88:547-552); and the octopine synthase promoter (Fromm et al. (1989) Plant Cell 1: 977-984).

A variety of plant gene promoters that regulate gene expression in response to environmental, hormonal, chemical, developmental signals, and in a tissue-active manner can be used for expression of a TF sequence in plants. Choice of a promoter is based largely on the phenotype of interest and is determined by such factors as tissue (e.g., seed, fruit, root, pollen, vascular tissue, flower, carpel, etc.), inducibility (e.g., in response to wounding, heat, cold, drought, light, pathogens, etc.), timing, developmental stage, and the like. Numerous known promoters have been characterized and can favorably be employed to promote expression of a polynucleotide of the invention in a transgenic plant or cell of interest. For example, tissue specific promoters include: seed-specific promoters (such as the napin, phaseolin or DC3 promoter described in US Pat. No. 5,773,697), fruit-specific promoters that are active during fruit ripening (such as the *dru1* promoter (US Pat. No. 5,783,393), or the 2A11 promoter (US Pat. No. 4,943,674) and the tomato polygalacturonase promoter (Bird et al. (1988) Plant Mol Biol 11:651), root-specific promoters, such as those disclosed in US Patent Nos. 5,618,988, 5,837,848 and 5,905,186, pollen-active promoters such as PTA29, PTA26 and PTA13 (US Pat. No. 5,792,929), promoters active in vascular tissue (Ringli and Keller (1998) Plant Mol Biol 37:977-988), flower-specific (Kaiser et al. (1995) Plant Mol Biol 28:231-243), pollen (Baerson et al. (1994) Plant Mol Biol 26:1947-1959), carpels (Ohl et al. (1990) Plant Cell 2:837-848), pollen and ovules (Baerson et al. (1993) Plant Mol Biol 22:255-267), auxin-inducible promoters (such as that described in van der Kop et al. (1999) Plant Mol Biol 39:979-990 or Baumann et al. (1999) Plant Cell 11:323-334), cytokinin-inducible promoter (Guevara-Garcia (1998) Plant Mol Biol 38:743-753), promoters responsive to gibberellin (Shi et al. (1998) Plant Mol Biol 38:1053-1060, Willmott et al. (1998) 38:817-825) and the like. Additional promoters are those that elicit expression in response to heat (Ainley et al. (1993) Plant Mol Biol 22: 13-23), light (e.g., the pea *rbcS-3A* promoter, Kuhlemeier et al. (1989) Plant Cell 1:471, and the maize *rbcS* promoter, Schaffner and Sheen (1991) Plant Cell 3: 997); wounding (e.g., *wun1*, Siebertz et al. (1989) Plant Cell 1: 961); pathogens (such as the PR-1

promoter described in Buchel et al. (1999) Plant Mol. Biol. 40:387-396, and the PDF1.2 promoter described in Manners et al. (1998) Plant Mol. Biol. 38:1071-80), and chemicals such as methyl jasmonate or salicylic acid (Gatz et al. (1997) Plant Mol Biol 48: 89-108). In addition, the timing of the expression can be controlled by using promoters such as those acting at senescence (An and Amazon (1995) Science 270: 1986-1988); or late seed development (Odell et al. (1994) Plant Physiol 106:447-458).

Plant expression vectors can also include RNA processing signals that can be positioned within, upstream or downstream of the coding sequence. In addition, the expression vectors can include additional regulatory sequences from the 3'-untranslated region of plant genes, e.g., a 3' terminator region to increase mRNA stability of the mRNA, such as the PI-II terminator region of potato or the octopine or nopaline synthase 3' terminator regions.

Additional Expression Elements

Specific initiation signals can aid in efficient translation of coding sequences. These signals can include, e.g., the ATG initiation codon and adjacent sequences. In cases where a coding sequence, its initiation codon and upstream sequences are inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only coding sequence (e.g., a mature protein coding sequence), or a portion thereof, is inserted, exogenous transcriptional control signals including the ATG initiation codon can be separately provided. The initiation codon is provided in the correct reading frame to facilitate transcription. Exogenous transcriptional elements and initiation codons can be of various origins, both natural and synthetic. The efficiency of expression can be enhanced by the inclusion of enhancers appropriate to the cell system in use.

Expression Hosts

The present invention also relates to host cells which are transduced with vectors of the invention, and the production of polypeptides of the invention (including fragments thereof) by recombinant techniques. Host cells are genetically engineered (i.e., nucleic acids are introduced, e.g., transduced, transformed or transfected) with the vectors of this invention, which may be, for example, a cloning vector or an expression vector comprising the relevant nucleic acids herein. The

vector is optionally a plasmid, a viral particle, a phage, a naked nucleic acid, *etc.* The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants, or amplifying the relevant gene. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to those skilled in the art and in the references cited herein, including, Sambrook and Ausubel.

The host cell can be a eukaryotic cell, such as a yeast cell, or a plant cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Plant protoplasts are also suitable for some applications. For example, the DNA fragments are introduced into plant tissues, cultured plant cells or plant protoplasts by standard methods including electroporation (Fromm et al., (1985) Proc. Natl. Acad. Sci. USA 82, 5824, infection by viral vectors such as cauliflower mosaic virus (CaMV) (Hohn et al., (1982) Molecular Biology of Plant Tumors, (Academic Press, New York) pp. 549-560; US 4,407,956), high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface (Klein et al., (1987) Nature 327, 70-73), use of pollen as vector (WO 85/01856), or use of *Agrobacterium tumefaciens* or *A. rhizogenes* carrying a T-DNA plasmid in which DNA fragments are cloned. The T-DNA plasmid is transmitted to plant cells upon infection by *Agrobacterium tumefaciens*, and a portion is stably integrated into the plant genome (Horsch et al. (1984) Science 233:496-498; Fraley et al. (1983) Proc. Natl. Acad. Sci. USA 80, 4803).

The cell can include a nucleic acid of the invention which encodes a polypeptide, wherein the cells expresses a polypeptide of the invention. The cell can also include vector sequences, or the like. Furthermore, cells and transgenic plants that include any polypeptide or nucleic acid above or throughout this specification, e.g., produced by transduction of a vector of the invention, are an additional feature of the invention.

For long-term, high-yield production of recombinant proteins, stable expression can be used. Host cells transformed with a nucleotide sequence encoding a polypeptide of the invention are optionally cultured under conditions suitable for the

expression and recovery of the encoded protein from cell culture. The protein or fragment thereof produced by a recombinant cell may be secreted, membrane-bound, or contained intracellularly, depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides encoding mature proteins of the invention can be designed with signal sequences which direct secretion of the mature polypeptides through a prokaryotic or eukaryotic cell membrane.

XI. Modified Amino Acid Residues

Polypeptides of the invention may contain one or more modified amino acid residues. The presence of modified amino acids may be advantageous in, for example, increasing polypeptide half-life, reducing polypeptide antigenicity or toxicity, increasing polypeptide storage stability, or the like. Amino acid residue(s) are modified, for example, co-translationally or post-translationally during recombinant production or modified by synthetic or chemical means.

Non-limiting examples of a modified amino acid residue include incorporation or other use of acetylated amino acids, glycosylated amino acids, sulfated amino acids, prenylated (e.g., farnesylated, geranylgeranylated) amino acids, PEG modified (e.g., "PEGylated") amino acids, biotinylated amino acids, carboxylated amino acids, phosphorylated amino acids, etc. References adequate to guide one of skill in the modification of amino acid residues are replete throughout the literature.

The modified amino acid residues may prevent or increase affinity of the polypeptide for another molecule, including, but not limited to, polynucleotide, proteins, carbohydrates, lipids and lipid derivatives, and other organic or synthetic compounds.

XII. Identification of Additional Factors

A transcription factor provided by the present invention can also be used to identify additional endogenous or exogenous molecules that can affect a phenotype or trait of interest. On the one hand, such molecules include organic (small or large molecules) and/or inorganic compounds that affect expression of (i.e., regulate) a particular transcription factor. Alternatively, such molecules include endogenous

molecules that are acted upon either at a transcriptional level by a transcription factor of the invention to modify a phenotype as desired. For example, the transcription factors can be employed to identify one or more downstream gene with which is subject to a regulatory effect of the transcription factor. In one approach, a transcription factor or transcription factor homologue of the invention is expressed in a host cell, e.g., a transgenic plant cell, tissue or explant, and expression products, either RNA or protein, of likely or random targets are monitored, e.g., by hybridization to a microarray of nucleic acid probes corresponding to genes expressed in a tissue or cell type of interest, by two-dimensional gel electrophoresis of protein products, or by any other method known in the art for assessing expression of gene products at the level of RNA or protein. Alternatively, a transcription factor of the invention can be used to identify promoter sequences (i.e., binding sites) involved in the regulation of a downstream target. After identifying a promoter sequence, interactions between the transcription factor and the promoter sequence can be modified by changing specific nucleotides in the promoter sequence or specific amino acids in the transcription factor that interact with the promoter sequence to alter a plant trait. Typically, transcription factor DNA-binding sites are identified by gel shift assays. After identifying the promoter regions, the promoter region sequences can be employed in double-stranded DNA arrays to identify molecules that affect the interactions of the transcription factors with their promoters (Bulyk et al. (1999) Nature Biotechnology 17:573-577).

The identified transcription factors are also useful to identify proteins that modify the activity of the transcription factor. Such modification can occur by covalent modification, such as by phosphorylation, or by protein-protein (homo or-heteropolymer) interactions. Any method suitable for detecting protein-protein interactions can be employed. Among the methods that can be employed are co-immunoprecipitation, cross-linking and co-purification through gradients or chromatographic columns, and the two-hybrid yeast system.

The two-hybrid system detects protein interactions in vivo and is described in Chien et al. ((1991), Proc. Natl. Acad. Sci. USA 88:9578-9582) and is commercially available from Clontech (Palo Alto, Calif.). In such a system, plasmids are constructed that encode two hybrid proteins: one consists of the DNA-binding domain

of a transcription activator protein fused to the TF polypeptide and the other consists of the transcription activator protein's activation domain fused to an unknown protein that is encoded by a cDNA that has been recombined into the plasmid as part of a cDNA library. The DNA-binding domain fusion plasmid and the cDNA library are transformed into a strain of the yeast *Saccharomyces cerevisiae* that contains a reporter gene (e.g., lacZ) whose regulatory region contains the transcription activator's binding site. Either hybrid protein alone cannot activate transcription of the reporter gene. Interaction of the two hybrid proteins reconstitutes the functional activator protein and results in expression of the reporter gene, which is detected by an assay for the reporter gene product. Then, the library plasmids responsible for reporter gene expression are isolated and sequenced to identify the proteins encoded by the library plasmids. After identifying proteins that interact with the transcription factors, assays for compounds that interfere with the TF protein-protein interactions can be preformed.

XIII. Identification of Modulators

In addition to the intracellular molecules described above, extracellular molecules that alter activity or expression of a transcription factor, either directly or indirectly, can be identified. For example, the methods can entail first placing a candidate molecule in contact with a plant or plant cell. The molecule can be introduced by topical administration, such as spraying or soaking of a plant, and then the molecule's effect on the expression or activity of the TF polypeptide or the expression of the polynucleotide monitored. Changes in the expression of the TF polypeptide can be monitored by use of polyclonal or monoclonal antibodies, gel electrophoresis or the like. Changes in the expression of the corresponding polynucleotide sequence can be detected by use of microarrays, Northern, quantitative PCR, or any other technique for monitoring changes in mRNA expression. These techniques are exemplified in Ausubel et al. (eds) Current Protocols in Molecular Biology, John Wiley & Sons (1998, and supplements through 2001). Such changes in the expression levels can be correlated with modified plant traits and thus identified molecules can be useful for soaking or spraying on fruit, vegetable and grain crops to modify traits in plants.

Essentially any available composition can be tested for modulatory activity of expression or activity of any nucleic acid or polypeptide herein. Thus, available libraries of compounds such as chemicals, polypeptides, nucleic acids and the like can be tested for modulatory activity. Often, potential modulator compounds can be dissolved in aqueous or organic (e.g., DMSO-based) solutions for easy delivery to the cell or plant of interest in which the activity of the modulator is to be tested. Optionally, the assays are designed to screen large modulator composition libraries by automating the assay steps and providing compounds from any convenient source to assays, which are typically run in parallel (e.g., in microtiter formats on microtiter plates in robotic assays).

In one embodiment, high throughput screening methods involve providing a combinatorial library containing a large number of potential compounds (potential modulator compounds). Such "combinatorial chemical libraries" are then screened in one or more assays, as described herein, to identify those library members (particular chemical species or subclasses) that display a desired characteristic activity. The compounds thus identified can serve as target compounds.

A combinatorial chemical library can be, e.g., a collection of diverse chemical compounds generated by chemical synthesis or biological synthesis. For example, a combinatorial chemical library such as a polypeptide library is formed by combining a set of chemical building blocks (e.g., in one example, amino acids) in every possible way for a given compound length (i.e., the number of amino acids in a polypeptide compound of a set length). Exemplary libraries include peptide libraries, nucleic acid libraries, antibody libraries (see, e.g., Vaughn et al. (1996) Nature Biotechnology, 14(3):309-314 and PCT/US96/10287), carbohydrate libraries (see, e.g., Liang et al. Science (1996) 274:1520-1522 and U.S. Patent 5,593,853), peptide nucleic acid libraries (see, e.g., U.S. Patent 5,539,083), and small organic molecule libraries (see, e.g., benzodiazepines, Baum C&EN Jan 18, page 33 (1993); isoprenoids, U.S. Patent 5,569,588; thiazolidinones and metathiazanones, U.S. Patent 5,549,974; pyrrolidines, U.S. Patents 5,525,735 and 5,519,134; morpholino compounds, U.S. Patent 5,506,337) and the like.

Preparation and screening of combinatorial or other libraries is well known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, peptide libraries (see, e.g., U.S. Patent 5,010,175; Furka, (1991) Int. J. Pept. Prot. Res. 37:487-493; and Houghton et al. (1991) Nature 354:84-88). Other chemistries for generating chemical diversity libraries can also be used.

In addition, as noted, compound screening equipment for high-throughput screening is generally available, e.g., using any of a number of well known robotic systems that have also been developed for solution phase chemistries useful in assay systems. These systems include automated workstations including an automated synthesis apparatus and robotic systems utilizing robotic arms. Any of the above devices are suitable for use with the present invention, e.g., for high-throughput screening of potential modulators. The nature and implementation of modifications to these devices (if any) so that they can operate as discussed herein will be apparent to persons skilled in the relevant art.

Indeed, entire high throughput screening systems are commercially available. These systems typically automate entire procedures including all sample and reagent pipetting, liquid dispensing, timed incubations, and final readings of the microplate in detector(s) appropriate for the assay. These configurable systems provide high throughput and rapid start up as well as a high degree of flexibility and customization. Similarly, microfluidic implementations of screening are also commercially available.

The manufacturers of such systems provide detailed protocols the various high throughput. Thus, for example, Zymark Corp. provides technical bulletins describing screening systems for detecting the modulation of gene transcription, ligand binding, and the like. The integrated systems herein, in addition to providing for sequence alignment and, optionally, synthesis of relevant nucleic acids, can include such screening apparatus to identify modulators that have an effect on one or more polynucleotides or polypeptides according to the present invention.

In some assays it is desirable to have positive controls to ensure that the components of the assays are working properly. At least two types of positive controls are appropriate. That is, known transcriptional activators or inhibitors can be

incubated with cells/plants/ etc. in one sample of the assay, and the resulting increase/decrease in transcription can be detected by measuring the resulting increase in RNA/ protein expression, etc., according to the methods herein. It will be appreciated that modulators can also be combined with transcriptional activators or inhibitors to find modulators that inhibit transcriptional activation or transcriptional repression. Either expression of the nucleic acids and proteins herein or any additional nucleic acids or proteins activated by the nucleic acids or proteins herein, or both, can be monitored.

In an embodiment, the invention provides a method for identifying compositions that modulate the activity or expression of a polynucleotide or polypeptide of the invention. For example, a test compound, whether a small or large molecule, is placed in contact with a cell, plant (or plant tissue or explant), or composition comprising the polynucleotide or polypeptide of interest and a resulting effect on the cell, plant, (or tissue or explant) or composition is evaluated by monitoring, either directly or indirectly, one or more of: expression level of the polynucleotide or polypeptide, activity (or modulation of the activity) of the polynucleotide or polypeptide. In some cases, an alteration in a plant phenotype can be detected following contact of a plant (or plant cell, or tissue or explant) with the putative modulator, e.g., by modulation of expression or activity of a polynucleotide or polypeptide of the invention. Modulation of expression or activity of a polynucleotide or polypeptide of the invention may also be caused by molecular elements in a signal transduction second messenger pathway and such modulation can affect similar elements in the same or another signal transduction second messenger pathway.

XIV. Subsequences

Also contemplated are uses of polynucleotides, also referred to herein as oligonucleotides, typically having at least 12 bases, preferably at least 15, more preferably at least 20, 30, or 50 bases, which hybridize under at least highly stringent (or ultra-high stringent or ultra-ultra-high stringent conditions) conditions to a polynucleotide sequence described above. The polynucleotides may be used as probes, primers, sense and antisense agents, and the like, according to methods as noted *supra*.

Subsequences of the polynucleotides of the invention, including polynucleotide fragments and oligonucleotides are useful as nucleic acid probes and primers. An oligonucleotide suitable for use as a probe or primer is at least about 15 nucleotides in length, more often at least about 18 nucleotides, often at least about 21 nucleotides, frequently at least about 30 nucleotides, or about 40 nucleotides, or more in length. A nucleic acid probe is useful in hybridization protocols, e.g., to identify additional polypeptide homologues of the invention, including protocols for microarray experiments. Primers can be annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand, and then extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods. See Sambrook and Ausubel, *supra*.

In addition, the invention includes an isolated or recombinant polypeptide including a subsequence of at least about 15 contiguous amino acids encoded by the recombinant or isolated polynucleotides of the invention. For example, such polypeptides, or domains or fragments thereof, can be used as immunogens, e.g., to produce antibodies specific for the polypeptide sequence, or as probes for detecting a sequence of interest. A subsequence can range in size from about 15 amino acids in length up to and including the full length of the polypeptide.

To be encompassed by the present invention, an expressed polypeptide which comprises such a polypeptide subsequence performs at least one biological function of the intact polypeptide in substantially the same manner, or to a similar extent, as does the intact polypeptide. For example, a polypeptide fragment can comprise a recognizable structural motif or functional domain such as a DNA binding domain that binds to a specific DNA promoter region, an activation domain or a domain for protein-protein interactions.

XV. Production of Transgenic Plants

Modification of Traits

The polynucleotides of the invention are favorably employed to produce transgenic plants with various traits, or characteristics, that have been modified in a desirable manner, e.g., to improve the seed characteristics of a plant. For example, alteration of expression levels or patterns (e.g., spatial or temporal expression patterns) of one or more of the transcription factors (or transcription factor homologues) of the invention, as compared with the levels of the same protein found in a wild type plant, can be used to modify a plant's traits. An illustrative example of trait modification, improved characteristics, by altering expression levels of a particular transcription factor is described further in the Examples and the Sequence Listing.

Arabidopsis as a model system

Arabidopsis thaliana is the object of rapidly growing attention as a model for genetics and metabolism in plants. *Arabidopsis* has a small genome, and well documented studies are available. It is easy to grow in large numbers and mutants defining important genetically controlled mechanisms are either available, or can readily be obtained. Various methods to introduce and express isolated homologous genes are available (see Koncz, et al., eds. *Methods in Arabidopsis Research*. et al. (1992), World Scientific, New Jersey, New Jersey, in "Preface"). Because of its small size, short life cycle, obligate autogamy and high fertility, *Arabidopsis* is also a choice organism for the isolation of mutants and studies in morphogenetic and development pathways, and control of these pathways by transcription factors (Koncz, *supra*, p. 72). A number of studies introducing transcription factors into *A. thaliana* have demonstrated the utility of this plant for understanding the mechanisms of gene regulation and trait alteration in plants. See, for example, Koncz, *supra*, and U.S. Patent Number 6,417,428).

Arabidopsis genes in transgenic plants.

Expression of genes which encode transcription factors modify expression of endogenous genes, polynucleotides, and proteins are well known in the art. In addition, transgenic plants comprising isolated polynucleotides encoding transcription factors may also modify expression of endogenous genes, polynucleotides, and

proteins. Examples include Peng et al. (1997, Genes and Development 11:3194-3205) and Peng et al. (1999, Nature, 400:256-261). In addition, many others have demonstrated that an *Arabidopsis* transcription factor expressed in an exogenous plant species elicits the same or very similar phenotypic response. See, for example, Fu et al. (2001, Plant Cell 13:1791-1802); Nandi et al. (2000, Curr. Biol. 10:215-218); Coupland (1995, Nature 377:482-483); and Weigel and Nilsson (1995, Nature 377:482-500).

Homologous genes introduced into transgenic plants.

Homologous genes that may be derived from any plant, or from any source whether natural, synthetic, semi-synthetic or recombinant, and that share significant sequence identity or similarity to those provided by the present invention, may be introduced into plants, for example, crop plants, to confer desirable or improved traits. Consequently, transgenic plants may be produced that comprise a recombinant expression vector or cassette with a promoter operably linked to one or more sequences homologous to presently disclosed sequences. The promoter may be, for example, a plant or viral promoter.

The invention thus provides for methods for preparing transgenic plants, and for modifying plant traits. These methods include introducing into a plant a recombinant expression vector or cassette comprising a functional promoter operably linked to one or more sequences homologous to presently disclosed sequences. Plants and kits for producing these plants that result from the application of these methods are also encompassed by the present invention.

The complete descriptions of the traits associated with each polynucleotide of the invention is fully disclosed in Table 4, Table 5, and Table 6.

Table 4

Polynucleotide SEQ ID NO:	GID No.	Trait	Category	Family	Comment	Polypeptide SEQ ID NO:	Conserved domains
1	G1275	Architecture; size	Dev and morph	WRKY	Reduced apical dominance; small plant	2	(113-169)
3	G1411	Architecture; light response; size; seed protein content	Dev and morph biochemistry	AP2	Loss of apical dominance	4	(87-154)
5	G1488	Architecture; flower; morphology; other	Dev and morph biochemistry	GATA/Zn	Reduced apical dominance, shorter stems; constitutive photomorphogenesis; reduced size; altered seed protein content	6	(221-246)
7	G1499	Architecture; flower; morphology; other; seed oil	Dev and morph biochemistry	HLH/MYC	Altered plant architecture; altered floral organ identity and development; dark green color	8	(118-181)
9	G1543	Architecture; morphology; other; fertility	Dev and morph biochemistry	HB	Altered plant architecture; altered carpel shape; dark green color; decreased seed oil	10	(135-195)
11	G1635	Architecture; light response; seed oil and protein content	Dev and morph biochemistry	MYB-related	Reduced apical dominance; pale green, smaller plants; reduced fertility	12	(44-104)
13	G1794	Architecture; size	Dev and morph biochemistry	AP2	Altered plant architecture; constitutive photomorphogenesis; altered seed oil and protein content	14	(182-248)
15	G1839	Architecture	Dev and morph	AP2	Altered plant architecture; reduced size	16	(118-184)
17	G2108	Architecture	Dev and morph	AP2	Altered inflorescence structure	18	(18-85)
19	G2291	Architecture; flowering time	Dev and morph flowering time	AP2	Altered plant architecture; late flowering	20	(TBD)
21	G2452	Architecture; leaf	Dev and morph	MYB-related	Reduced apical dominance; pale green color	22	(27-213)
23	G2509	Architecture; seed oil and protein content	Dev and morph biochemistry	AP2	Reduced apical dominance; altered seed oil and protein content	24	(89-156)
25	G390	Architecture	Dev and morph	HB	Altered shoot development	26	(18-81)
27	G391	Architecture	Dev and morph	HB	Altered shoot development	28	(25-85)
29	G438	Architecture; stem	Dev and morph	HB	Reduced branching; reduced lignin	30	(22-85)

Table 4

31	G47	Architecture; stem; flowering time; altered seed oil content	Dev and morph; flowering time; seed biochemistry	AP2	Altered architecture and inflorescence development, structure of vascular tissues; late flowering; altered seed oil content	32	(11-80)
33	G559	Architecture; fertility	Dev and morph; bZIP		Loss of apical dominance; reduced fertility	34	(203-264)
35	G568	Architecture; flowering time	Dev and morph; bZIP		Altered branching; late flowering	36	(215-265)
37	G580	Architecture; flower	Dev and morph; bZIP		Altered inflorescences; altered flower development	38	(162-218)
39	G615	Architecture; fertility	Dev and morph; TEO		Altered plant architecture; little or no pollen production, poor filament elongation	40	(88-147)
41	G732	Architecture; flower; seed oil and protein	Dev and morph; biochemistry	bZIP	Reduced apical dominance; abnormal flowers; altered seed oil and protein content	42	(31-91)
43	G988	Architecture; fertility; flower; stem; seed oil and protein content	Dev and morph; biochemistry	SCR	Reduced lateral branching; reduced fertility; enlarged floral organs, short pedicels; thicker stem, altered distribution of vascular bundles; altered seed oil and protein content	44	(178-195)
45	G1519	Embryo lethal	Dev and morph; RING/C3HC4		Embryo lethal	46	(327-364)
47	G374	Embryo lethal	Dev and morph; Z-ZPF		Embryo lethal	48	(35-67, 245-277)
49	G877	Embryo lethal	Dev and morph; WRKY		Embryo lethal	50	(272-328, 487-603)
51	G1000	Fertility; size; flower; stem	Dev and morph	MYB-(R1)R2R3	Reduced fertility; small plant; reduced or absent petals and sepals; reduced inflorescence, stem elongation	52	(14-117)
53	G1067	Fertility; leaf; size	Dev and morph; AT-hook		Reduced fertility; altered leaf shape; small plant	54	(86-93)
55	G1075	Fertility; flower; leaf; size	Dev and morph; AT-hook		Reduced fertility; reduced or absent petals, sepals and stamens; altered leaf shape; small plant	56	(78-85)
57	G1266	Fertility; size	Dev and morph; AP2		Reduced fertility; small plant	58	(79-147)
59	G1311	Fertility; size	Dev and morph; MYB-(R1)R2R3		Reduced fertility; small plant	60	(11-112)
61	G1321	Fertility; flower	Dev and morph; MYB-(R1)R2R3		Poor fertility; altered flower morphology	62	(4-106)
63	G1326	Fertility; flower; size	Dev and morph; MYB-(R1)R2R3		Reduced fertility; petals and sepals are smaller; small plant	64	(18-121)
65	G1367	Fertility; size	Dev and morph; AT-hook		Reduced fertility; reduced size	66	(179-201, 262-285, 298-319, 335-357)
67	G1386	Fertility; size; seed oil and protein content	Dev and morph; biochemistry	AP2	Reduced fertility; reduced size; altered seed oil and protein content	68	(TBD)

Table 4

69	G1421	Fertility; size; seed oil content	Dev and morph; seed biochemistry	AP2	Reduced fertility; small plant; altered seed oil content	70	(74-151)
71	G1453	Fertility; morphology; other	Dev and morph	NAC	Reduced fertility; altered inflorescence development	72	(13-160)
73	G1560	Fertility; flower; size	Dev and morph	HS	Reduced fertility; altered flower development; reduced size	74	(62-151)
75	G1594	Fertility; leaf; seed	Dev and morph	HB	Reduced fertility; altered leaf shape and development; large pale seed	76	(343-308)
77	G1750	Fertility; size; seed oil content	Dev and morph; seed biochemistry	AP2	Reduced fertility; reduced size; increased seed oil content	78	(107-173)
79	G1947	Fertility; flower; seed protein content	Dev and morph; seed biochemistry	HS	Reduced fertility; extended period of flowering; altered seed protein content	80	(37-120)
81	G2011	Fertility; size; seed oil and protein content	Dev and morph; seed biochemistry	HS	Reduced fertility; reduced size; altered seed oil and protein content	82	(56-147)
83	G2094	Fertility; leaf; size	Dev and morph	GATA/Zn	Reduced fertility; altered leaf development; reduced size	84	(43-68)
85	G2113	Fertility; leaf; seed protein content	Dev and morph; seed biochemistry	AP2	Reduced fertility; long petioles, altered orientation; altered seed protein content	86	(TBD)
87	G2115	Fertility; size	Dev and morph	AP2	Reduced fertility; reduced size	88	(46-115)
89	G2130	Fertility; size; senescence	Dev and morph	AP2	Reduced fertility; reduced size; early senescence	90	(93-160)
91	G2147	Fertility; size	Dev and morph	HLH/MYC	Reduced fertility; reduced size	92	(160-234)
93	G2156	Fertility; size; seed protein content	Dev and morph; seed biochemistry	AT-hook	Reduced fertility; reduced size; altered seed protein content	94	(66-86)
95	G2294	Fertility; size	Dev and morph	AP2	Reduced fertility; reduced size	96	(32-102)
97	G2510	Fertility; size	Dev and morph	AP2	Reduced fertility; reduced size	98	(41-108)
99	G2893	Fertility; flower; size	Dev and morph	MYB-(R1)/R2R3	Reduced fertility; altered flower development; reduced size	100	(19-120)
101	G340	Fertility; size	Dev and morph	Z-C3H	Reduced fertility, size	102	(37-154)
103	G39	Fertility; size	Dev and morph	AP2	Reduced fertility, small plant	104	(24-90)
105	G439	Fertility; size	Dev and morph	AP2	Reduced fertility, small plant	106	(110-177)
107	G470	Fertility	Dev and morph	ARF	Short stamen filaments	108	(61-393)

Table 4

109	G652	Fertility; seed; flower; size; seed oil content	Dev and morph; seed biochemistry	Z-CLDSH	Reduced fertility; irregular shaped seed; altered flower development; reduced size, slow growth; altered seed oil content	110	(28-49, 137-151, 182-196)
111	G671	Fertility; flower; leaf; size; stem	Dev and morph	MYB-(R1)R2R3	Reduced fertility; reduced petal abscission; altered leaf shape; small plant; altered inflorescence stem structure	112	(15-115)
113	G779	Fertility; flower	Dev and morph	HLH/MYC	Reduced fertility, homeotic transformations	114	(126-182)
115	G962	Fertility; size	Dev and morph	NAC	Reduced fertility; small plant	116	(53-175)
117	G977	Fertility; leaf; morphology; other; size	Dev and morph	AP2	Reduced fertility; altered leaf shape; dark green; small plant	118	(5-72)
119	G1063	Flower; leaf; inflorescence; seed oil and protein content	Dev and morph; seed biochemistry	HLH/MYC	Altered flower development; ectopic carpel tissue; altered leaf shape, dark green color; altered inflorescence development; altered seed oil and protein content	120	(131-182)
121	G1140	Flower	Dev and morph	MADS	Altered flower development	122	(2-57)
123	G1425	Flower	Dev and morph	NAC	Altered flower and inflorescence development	124	(20-173)
125	G1449	Flower	Dev and morph	IAA	Altered flower structure	126	(48-53, 74-107, 122-152)
127	G1897	Flower; leaf; seed protein content	Dev and morph; seed biochemistry	Z-Dof	Altered flower development; altered leaf development; altered seed protein content	128	(34-62)
129	G2143	Flower; leaf; inflorescence	Dev and morph	HLH/MYC	Altered flower development, ectopic carpel tissue; altered leaf shape, dark green color; altered inflorescence development	130	(128-179)
131	G2535	Flower; seed protein content	Dev and morph; seed biochemistry	NAC	Altered flower development; altered seed protein content	132	(11-114)
133	G2557	Flower; leaf	Dev and morph	HLH/MYC	Altered flower development, ectopic carpel tissue; altered leaf shape, dark green color	134	(278-328)
135	G259	Flower; leaf	Dev and morph	HS	Altered flower development; altered leaf development	136	(27-131)
137	G353	Flower; leaf; size; seed protein content	Dev and morph; seed biochemistry	Z-C2H2	Short pedicels, downward pointing siliques; altered leaf development; reduced size; altered seed protein content	138	(41-61, 84-104)
139	G354	Flower; light response; size	Dev and morph	Z-C2H2	Short pedicels, downward pointing siliques; constitutive morphogenesis; reduced size	140	(42-62, 88-109)
141	G538	Flower; morphology; other	Dev and morph	TH	Altered flower development; multiple developmental defects	142	(119-206)

Table 4

143	G869	Flower; morphology; other; seed oil	Dev and morph; seed biochemistry	AP2	Abnormal anther development; small and spindly plant; altered seed fatty acids	144	(109-177)
145	G1645	Inflorescence; leaf	Dev and morph	MYB-(R1)R2R3	Altered inflorescence structure; altered leaf development	146	(90-210)
147	G1038	Leaf	Dev and morph	GARP	Altered leaf shape	148	(198-247)
149	G1073	Leaf; size; flowering time	Dev and morph; flowering time	AT-hook	Serrated leaves; increased plant size; flowering appears to be slightly delayed	150	(33-42, 78-175)
151	G1146	Leaf	Dev and morph	PAZ	Altered leaf development	152	(886-896)
153	G1267	Leaf; size	Dev and morph	WRKY	Dark green shiny leaves; small plant	154	(70-127)
155	G1269	Leaf	Dev and morph	MYB-related	Long petioles, upturned leaves	156	(27-83)
157	G1452	Leaf; trichome; flowering time	Dev and morph; flowering time	NAC	Altered leaf shape, dark green color; reduced trichome density; late flowering	158	(30-177)
159	G1494	Leaf; size; light response; seed	Dev and morph	HLH/MYC	Pale green leaves, altered leaf shape; reduced size; long hypocotyls; large, pale seeds	160	(261-311)
161	G1548	Leaf	Dev and morph	HB	Altered leaf development	162	(17-77)
163	G1574	Leaf	Dev and morph	SWI/SNF	Altered leaf development	164	(28-350)
165	G1586	Leaf; size	Dev and morph	HB	Narrow leaves; small plants	166	(21-81)
167	G1786	Leaf; light response; size	Dev and morph	MYB-(R1)R2R3	Dark green; small leaves with short petioles; photomorphogenesis in the dark; small plant	168	(TBD)
169	G1792	Leaf; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Dark green, shiny leaves; altered seed oil and protein content	170	(17-85)
171	G1865	Leaf; seed oil and protein content	Dev and morph; seed biochemistry	GRF-like	Altered leaf development; altered seed oil and protein content	172	(124-149)
173	G1886	Leaf; size	Dev and morph	Z-Dof	Chlorotic patches in leaves; reduced size	174	(17-59)
175	G1933	Leaf; size; seed protein content	Dev and morph; seed biochemistry	WRKY	Altered leaf development; reduced size; altered seed protein content	176	(205-263, 344-404)
177	G2059	Leaf; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Smaller, curled leaves; altered seed oil, protein content	178	(184-254)
179	G2105	Leaf; seed oil and protein content	Dev and morph	TH	Alterations in leaf surface; large, pale seeds	180	(100-153)
181	G2117	Leaf; seed oil and protein content	Dev and morph; seed biochemistry	bZIP	Small, dark green leaves; altered seed oil and protein content	182	(46-106)

Table 4

183	G2124	Leaf; seed protein content	Dev and morph; seed biochemistry	TEO	Altered leaf development; altered seed protein content	184	(75-132)
185	G2140	Leaf; root	Dev and morph	HLH/MYC	Altered leaf development; short roots	186	(167-242)
187	G2144	Leaf; light response; size; seed oil content	Dev and morph; seed biochemistry	HLH/MYC	Pale green leaves, altered leaf shape; long hypocotyls; reduced size; altered seed oil content	188	(203-283)
189	G2431	Leaf	Dev and morph	GARP	Dark green leaves; reduced size	190	(38-88)
191	G2465	Morphology: other; leaf	Dev and morph	GARP	Slowed development; altered leaf color and shape	192	(219-269)
193	G2583	Leaf; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Glossy, shiny leaves; altered seed oil and protein content	194	(4-71)
195	G2724	Leaf	Dev and morph	MYB-(R1)R2R3	Dark green leaves	196	(7-113)
197	G377	Morphology: other	Dev and morph	RING/C3H2C3	Altered leaf development; slow growth	198	(85-128)
199	G428	Leaf	Dev and morph	HB	Altered leaf shape	200	(229-292)
201	G447	Leaf; morphology: other; size	Dev and morph	ARF	Dark green leaves; altered cotyledon shape; reduced size	202	(22-356)
203	G464	Leaf	Dev and morph	IAA	Altered leaf shape	204	(20-28, 71-82, 126-142, 187-224)
205	G557	Leaf; size	Dev and morph	bZIP	Dark green color; small plant	206	(90-150)
207	G577	Leaf	Dev and morph	BZIP2	Reduced size; increased anthocyanins	208	(TBD)
209	G674	Leaf; size	Dev and morph	MYB-(R1)R2R3	Dark green leaves, upwardly oriented; reduced size	210	(20-120)
211	G736	Leaf; flowering time	Dev and morph; flowering time	Z-Dof	Altered leaf shape; later flowering	212	(54-111)
213	G903	Leaf	Dev and morph	Z-C2H2	Altered leaf morphology	214	(68-92)
215	G917	Leaf; seed oil and protein content	Dev and morph; seed biochemistry	MADS	Altered leaf development; altered seed oil and protein content	216	(2-57)
217	G921	Leaf	Dev and morph	WRKY	Serrated leaves	218	(146-203)
219	G922	Leaf; size	Dev and morph	SCR	Altered development, dark green color; reduced size	220	(225-242)
221	G932	Leaf; size	Dev and morph	MYB-(R1)R2R3	Altered development, dark green color; reduced size	222	(12-118)
223	G599	Leaf; size	Dev and morph	DBP	Altered leaf shape; small plant	224	(187-219, 264-300)
225	G804	Leaf; size	Dev and morph	PCF	Altered leaf shape, small plant	226	(54-117)

Table 4

227	G1062	Light response; morphology; other; seed	Dev and morph	HLH/MYC	Constitutive photomorphogenesis; slow growth, altered seed shape	228	(308-359)
229	G1322	Light response; size	Dev and morph	MYB-(R1)R2R3	Photomorphogenesis in the dark; reduced size	230	(26-130)
231	G1331	Light response; morphology; other; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Constitutive photomorphogenesis; multiple developmental alterations; altered seed oil and protein content	232	(8-109)
233	G1521	Light response	Dev and morph	RING/C3HC4	Constitutive photomorphogenesis	234	(39-80)
235	G183	Light response; seed protein content	Dev and morph; seed biochemistry	WRKY	Constitutive photomorphogenesis; altered seed protein content	236	(307-363)
237	G2555	Light response	Dev and morph	HLH/MYC	Constitutive photomorphogenesis	238	(175-245)
239	G375	Light response	Dev and morph	Z-Dof	Upward pointing leaves	240	(75-103)
241	G1007	Morphology; other	Dev and morph	AP2	Multiple developmental alterations	242	(TBD)
243	G1010	Morphology; other	Dev and morph	ABI3/VP-1	Multiple developmental alterations	244	(33-122)
245	G1014	Morphology; other; trichome	Dev and morph	ABI3/VP-1	Multiple developmental defects; reduced trichomes	246	(90-172)
247	G1035	Morphology; other	Dev and morph	bZIP	Multiple developmental alterations	248	(39-91)
249	G1046	Morphology; other	Dev and morph	bZIP	Multiple developmental alterations	250	(79-138)
251	G1049	Morphology; other; seed protein content	Dev and morph; seed biochemistry	bZIP	Multiple developmental alterations; altered seed protein content	252	(77-132)
253	G1069	Morphology; other; seed oil content	Dev and morph; seed biochemistry	AT-hook	Multiple developmental alterations; altered seed oil content	254	(67-74)
255	G1070	Morphology; other	Dev and morph	AT-hook	Several developmental defects	256	(98-120)
257	G1076	Morphology; other	Dev and morph	AT-hook	Lethal when overexpressed	258	(82-89)
259	G1089	Morphology; other	Dev and morph	BZIPT2	Developmental defects at seedling stage	260	(425-500)
261	G1093	Morphology; other	Dev and morph	RING/C3H2C3	Multiple morphological alterations	262	(105-148)

Table 4

263	G1127	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	264	(103-110, 155-162)
265	G1131	Morphology: other; seed protein content	Dev and morph; seed biochemistry	HLH/MYC	Multiple developmental alterations; altered seed protein content	266	(173-220)
267	G1145	Morphology: other; seed oil and protein	Dev and morph; seed biochemistry	bZIP	Multiple developmental alterations; reduced seed size, altered seed shape; altered seed oil and protein content	268	(227-270)
269	G1229	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	HLH/MYC	Several developmental defects; altered seed oil and protein content	270	(102-160)
271	G1246	Morphology: other; seed protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Multiple developmental alterations; altered seed protein content	272	(27-139)
273	G1255	Morphology: other; seed	Dev and morph	Z-CO-like	Reduced apical dominance; increased seed size	274	(18-56)
275	G1304	Morphology: other	Dev and morph	MYB-(R1)R2R3	Lethal when overexpressed	276	(13-118)
277	G1318	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	278	(20-123)
279	G1320	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	280	(5-108)
281	G1330	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	282	(28-134)
283	G1352	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	284	(108-129, 167-188)
285	G1354	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	286	(TBD)
287	G1360	Morphology: other	Dev and morph	NAC	Lethal when overexpressed	288	(18-174)
289	G1364	Morphology: other	Dev and morph	CAAT	Lethal when overexpressed	290	(29-120)
291	G1379	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	292	(18-85)
293	G1384	Morphology: other	Dev and morph	AP2	Abnormal inflorescence and flower development	294	(TBD)
295	G1399	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	296	(86-93)

Table 4

297	G1415	Morphology: other	Dev and morph morph; seed biochemistry	AP2	Multiple developmental alterations	298	(TBD)
299	G1417	Morphology: other; seed oil		WRKY	Reduced seedling germination and vigor; increase in 18:2, decrease in 18:3	300	(239-296)
301	G1442	Morphology: other	Dev and morph	GRF-like	Multiple developmental alterations	302	(172-223)
303	G1454	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	NAC	Multiple developmental alterations; altered seed oil and protein content	304	(9-178)
305	G1459	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	306	(10-152)
307	G1460	Morphology: other; seed protein content	Dev and morph; seed biochemistry	NAC	Multiple developmental alterations; altered seed protein content	308	(TBD)
309	G147	Morphology: other	Dev and morph	MADS	Multiple developmental defects	310	(2-57)
311	G1471	Morphology: other; seed oil	Dev and morph; seed biochemistry	Z-C2H2	Multiple developmental alterations; increased seed oil content	312	(49-70)
313	G1475	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	314	(51-73)
315	G1477	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	316	(29-48)
317	G1487	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	GATA/Zn	Multiple developmental alterations; altered seed oil and protein content	318	(251-276)
319	G1492	Morphology: other	Dev and morph	GARP	Multiple developmental alterations	320	(34-83)
321	G1531	Morphology: other; seed; seed protein content	Dev and morph; seed biochemistry	RING/C3HC4	Multiple developmental alterations; pale seed; altered seed protein content	322	(41-77)
323	G1540	Morphology: other	Dev and morph	HB	Reduced cell differentiation in meristem	324	(35-98)
325	G1544	Morphology: other	Dev and morph	HB	Multiple developmental alterations	326	(64-124)

Table 4

327	G156	Morphology: other; seed	Dev and morph	MADS	Multiple developmental defects; seed color alteration	328	(2-57)
329	G1584	Morphology: other	Dev and morph	HB	Multiple developmental alterations	330	(TBD)
331	G1587	Morphology: other	Dev and morph	HB	Multiple developmental alterations	332	(61-121)
333	G1588	Morphology: other	Dev and morph	HB	Multiple developmental alterations	334	(66-124)
335	G1589	Morphology: other; seed protein content	Dev and morph; seed biochemistry	HB	Multiple developmental alterations; altered seed protein content	336	(384-448)
337	G160	Morphology: other	Dev and morph	MADS	Multiple developmental defects	338	(7-62)
339	G1636	Morphology: other	Dev and morph	MYB-related	Pale green, smaller plants	340	(100-165)
341	G1642	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	342	(TBD)
343	G1747	Morphology: other; seed protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Multiple developmental alterations; altered seed protein content	344	(11-114)
345	G1749	Morphology: other	Dev and morph	AP2	Multiple developmental alterations; formation of necrotic lesions	346	(84-155)
347	G1751	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	348	(TBD)
349	G1752	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	350	(83-151)
351	G1763	Morphology: other	Dev and morph	AP2	Lethal when overexpressed	352	(140-209)
353	G1766	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	354	(10-153)
355	G1767	Morphology: other; seed oil content	Dev and morph; seed biochemistry	SCR	Multiple developmental alterations; altered seed oil content	356	(255-272)
357	G1778	Morphology: other	Dev and morph	GATA/Zn	Lethal when overexpressed	358	(94-119)
359	G1789	Morphology: other; seed protein content	Dev and morph; seed biochemistry	MYB-related	Delayed development; altered seed protein content	360	(1-50)
361	G1790	Morphology: other	Dev and morph	MYB-(R1)R2R3	Lethal when overexpressed	362	(217-316)

Table 4

363	G1791	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	364	(TBD)
365	G1793	Morphology: other; seed oil	Dev and morph; biochemistry	AP2	Multiple developmental alterations; increased seed oil content	366	(179-255, 281-349)
367	G1795	Morphology: other; trichome	Dev and morph	AP2	Multiple developmental alterations; reduced trichomes	368	(12-80)
369	G1800	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	370	(TBD)
371	G1806	Morphology: other	Dev and morph	bZIP	Multiple developmental alterations	372	(165-225)
373	G1811	Morphology: other	Dev and morph	ABI3/V/P-1	Multiple developmental alterations	374	(TBD)
375	G182	Morphology: other	Dev and morph	WRKY	Multiple developmental alterations	376	(217-276)
377	G1835	Morphology: other	Dev and morph	GATA/Zn	Small, spindly plant	378	(224-296)
379	G1836	Morphology: other	Dev and morph	CAAT	Pale green	380	(30-164)
381	G1838	Morphology: other; seed oil content	Dev and morph; biochemistry	AP2	Multiple developmental alterations; increased seed oil content	382	(229-305, 330-400)
383	G1843	Morphology: other	Dev and morph	MADS	Multiple developmental alterations	384	(2-57)
385	G1853	Morphology: other	Dev and morph	AKR	Lethal when overexpressed	386	(entire protein)
387	G1855	Morphology: other	Dev and morph	AKR	Slow growth	388	(entire protein)
389	G187	Morphology: other	Dev and morph	WRKY	Variety of morphological alterations	390	(172-228)
391	G1881	Morphology: other	Dev and morph	Z-CO-like	Multiple developmental alterations	392	(5-28, 56-79)
393	G1882	Morphology: other	Dev and morph	Z-Dof	Lethal when overexpressed	394	(97-125)
395	G1883	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	396	(82-124)
397	G1884	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	398	(43-71)
399	G1891	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	400	(27-69)

Table 4

401	G1896	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	402	(43-85)
403	G1898	Morphology: other	Dev and morph	Z-Dof	Lethal when overexpressed	404	(31-59)
405	G1902	Morphology: other; seed oil content	Dev and morph; seed biochemistry	Z-Dof	Multiple developmental alterations; increased seed oil content	406	(31-59)
407	G1904	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	408	(53-95)
409	G1906	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	410	(19-47)
411	G1913	Morphology: other	Dev and morph	Z-Dof	Lethal when overexpressed	412	(27-55)
413	G1914	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	414	(195-216, 245-266)
415	G1925	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	416	(6-150)
417	G1929	Morphology: other	Dev and morph	Z-CO-like	Slow growth, delayed development	418	(31-53)
419	G1930	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	420	(59-124)
421	G195	Morphology: other	Dev and morph	WRKY	Multiple developmental defects	422	(183-239)
423	G1954	Morphology: other	Dev and morph	HLH/MYC	Lethal when overexpressed	424	(187-259)
425	G1958	Morphology: other; seed protein content	Dev and morph; seed biochemistry	GARP	Reduced size and root mass in plates.; altered seed protein content	426	(230-278)
427	G196	Morphology: other; seed protein content	Dev and morph; seed biochemistry	WRKY	Multiple developmental alterations; altered seed protein content	428	(223-283)
429	G1965	Morphology: other	Dev and morph	Z-Dof	Lethal when overexpressed	430	(27-55)
431	G1976	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	432	(219-323)
433	G2057	Morphology: other	Dev and morph	TEO	Multiple developmental alterations	434	(TBD)
435	G2107	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	436	(TBD)

Table 4

437	G211	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	438	(24-137)
439	G2133	Morphology: other; flowering time; seed protein content	Dev and morph; flowering time	AP2	Multiple developmental alterations; late flowering; altered seed protein content	440	(11-83)
441	G2134	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	442	(TBD)
443	G2151	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	AT-hook	Multiple developmental alterations; altered seed oil and protein content	444	(93-113, 124-144)
445	G2154	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	446	(97-119)
447	G2157	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	448	(82-102, 164-107)
449	G2181	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	450	(22-169)
451	G221	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	452	(21-125)
453	G2290	Morphology: other	Dev and morph	WRKY	Multiple developmental alterations	454	(147-205)
455	G2299	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	456	(48-115)
457	G2340	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Tissue necrosis; multiple developmental alterations; altered seed oil and protein content	458	(14-120)
459	G2346	Morphology: other	Dev and morph	SBP	Enlarged seedlings	460	(59-135)
461	G237	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	462	(11-113)
463	G2373	Morphology: other; seed protein content	Dev and morph; seed biochemistry	TH	Multiple developmental alterations; altered seed protein content	464	(290-350)
465	G2376	Morphology: other; seed oil protein	Dev and morph; seed biochemistry	TH	Seedling lethality; altered seed protein content	466	(79-178, 336-408)
467	G24	Morphology: other	Dev and morph	AP2	Reduced size and necrotic patches	468	(25-93)

Table 4

469	G2424	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	470	(107-219)
471	G2505	Morphology: other	Dev and morph	NAC	Lethal when overexpressed	472	(10-159)
473	G2512	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	474	(79-139)
475	G2513	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	476	(TBD)
477	G2519	Morphology: other	Dev and morph	HLH/MYC	Multiple developmental alterations	478	(1-65)
479	G2520	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	HLH/MYC	Multiple developmental alterations; altered seed oil and protein content	480	(135-206)
481	G2533	Morphology: other; seed protein content	Dev and morph; seed biochemistry	NAC	Multiple developmental alterations; altered seed protein content	482	(11-186)
483	G2534	Morphology: other	Dev and morph	NAC	Lethal when overexpressed	484	(10-157)
485	G2573	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Multiple developmental alterations; altered seed oil and protein content	486	(31-98)
487	G2589	Morphology: other	Dev and morph	MADS	Multiple developmental alterations	488	(2-57)
489	G2687	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	490	(51-120)
491	G27	Morphology: other	Dev and morph	AP2	Abnormal development, small	492	(37-104)
493	G2720	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Multiple developmental alterations; altered seed oil and protein content	494	(10-114) (172-192, 226-247, 256-276, 290-311, 245 366)
495	G2787	Morphology: other; seed oil content	Dev and morph; seed biochemistry	AT-hook	Multiple developmental alterations; altered seed oil content	496	
497	G2789	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	498	(53-73, 121-165)
499	G31	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	500	(TBD)

Table 4

501	G33	Morphology: other	Dev and morph	AP2	Multiple developmental defects	502	(50-117)
503	G342	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	GATA/Zn	Multiple developmental alterations; altered seed oil and protein content	504	(155-190)
505	G352	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	506	(99-119, 166-186)
507	G357	Morphology: other	Dev and morph	Z-C2H2	Developmental defect	508	(7-29)
509	G358	Morphology: other	Dev and morph	Z-C2H2	Lethal when overexpressed	510	(124-135, 188-210)
511	G360	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	512	(42-62)
513	G362	Size; Morphology: other; trichome; flowering time; seed protein content	Dev and morph; morph; flowering time; seed biochemistry	Z-C2H2	Reduced size; increased pigmentation in seed, embryos and other organs; ectopic trichome formation; increased trichome number, late flowering; altered protein content	514	(62-82)
515	G364	Morphology: other	Dev and morph	Z-C2H2	Developmental defect	516	(54-76)
517	G365	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	518	(70-90)
519	G367	Morphology: other	Dev and morph	Z-C2H2	Lethal when overexpressed	520	(63-84)
521	G373	Morphology: other	Dev and morph	RING/C3HC4	Multiple developmental alterations	522	(129-168)
523	G396	Morphology: other; size	Dev and morph	HB	Altered leaf coloration and shape, reduced fertility; small plant	524	(159-220)
525	G431	Morphology: other	Dev and morph	HB	Developmental defect, sterile	526	(286-335)
527	G479	Morphology: other	Dev and morph	SBP	Multiple developmental alterations	528	(70-149)
529	G546	Morphology: other	Dev and morph	RING/C3H2C3	Slow growth and development; increased anthocyanin pigmentation	530	(114-155)
531	G551	Morphology: other	Dev and morph	HB	Multiple developmental alterations	532	(73-133)
533	G578	Morphology: other	Dev and morph	bZIP	Lethal when overexpressed	534	(36-96)

Table 4

535	G596	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	536	(89-96)
537	G617	Morphology: other	Dev and morph	TEO	Multiple developmental alterations	538	(64-118)
539	G620	Morphology: other; seed protein content	Dev and morph; seed biochemistry	CAAT	Multiple developmental alterations; altered seed protein content	540	(20-118)
541	G625	Morphology: other	Dev and morph	AP2	Lethal when overexpressed	542	(52-119)
543	G658	Morphology: other	Dev and morph	MYB-(R1)R2R3	Developmental defect	544	(2-105)
545	G716	Morphology: other	Dev and morph	ARF	Multiple developmental defects	546	(24-355)
547	G725	Morphology: other	Dev and morph	GARP	Developmental defect	548	(39-87)
549	G727	Morphology: other	Dev and morph	GARP	Multiple morphological alterations	550	(226-269)
551	G740	Morphology: other	Dev and morph	Z-CLDSH	Slow growth	552	(24-42, 232-268)
553	G770	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	554	(19-162)
555	G858	Morphology: other	Dev and morph	MADS	Multiple developmental alterations	556	(2-57)
557	G865	Morphology: other; seed protein content	Dev and morph; seed biochemistry	AP2	Altered morphology; increased seed protein	558	(36-103)
559	G872	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	560	(18-85)
561	G904	Morphology: other	Dev and morph	RING/C3H2C3	Multiple developmental alterations	562	(117-158)
563	G910	Morphology: other; flowering time	Dev and morph; flowering time	Z-CO-like	Multiple developmental alterations; late flowering	564	(14-37, 77-103)
565	G912	Morphology: other; size; sugar sensing; flowering time	Dev and morph; sensing; flowering time	AP2	Dark green color; small plant; reduced cotyledon expansion in glucose; late flowering	566	(51-118)
567	G920	Morphology: other	Dev and morph	WRKY	Multiple developmental alterations	568	(152-211)

Table 4

569	G939	Morphology: other; size	Dev and morph Dev and morph; seed biochemistry	EIL	Pale seedlings on agar; reduced size	570	(97-106)
571	G963	Morphology: other; seed protein content	Dev and morph; seed biochemistry	NAC	Slowed growth rate; altered seed protein content	572	(TBD)
573	G979	Morphology: other; seed	Dev and morph	AP2	Several developmental defects; altered seed development, ripening and germination	574	(63-139, 165-233)
575	G987	Morphology: other	Dev and morph	SCR	Developmental defects	576	(428-432, 704-708)
577	G993	Morphology: other; seed protein content	Dev and morph; seed biochemistry	AP2	Multiple developmental alterations; altered seed protein content	578	(69-134)
579	G681	Morphology: other; leaf glucosinolates	Dev and morph; leaf biochemistry	MYB-(R1)R2R3	Multiple developmental alterations; overexpression results in an increase in M39480	580	(14-120)
581	G1482	Root	Dev and morph	Z-CO-like	Increased root growth	582	(5-63)
583	G225	Root; trichome	Dev and morph	MYB-related	Increased root hairs; glabrous, lack of trichomes	584	(39-76)
585	G226	Root; trichome; seed protein content	Dev and morph; seed biochemistry	MYB-related	Increased root hairs; glabrous, lack of trichomes;	586	(28-78)
587	G9	Root	Dev and morph	AP2	Increased seed protein	588	(62-127)
589	G1040	Seed	Dev and morph	GARP	Increased root mass	590	(109-158)
591	G2114	Seed	Dev and morph	AP2	Smaller and more rounded seeds	592	(221-297, 323-393)
593	G450	Seed; size; seed protein content	Dev and morph; seed biochemistry	IAA	Increased seed size; reduced plant size; altered seed protein content	594	(TBD)
595	G584	Seed	Dev and morph	HLH/MYC	Large seeds	596	(401-494)
597	G668	Seed	Dev and morph	MYB-(R1)R2R3	Reduced seed color	598	(13-113)
599	G1050	Senescence	Dev and morph	bZIP	Delayed senescence	600	(372-425)
601	G1463	Senescence	Dev and morph	NAC	Premature senescence	602	(9-156)
603	G1944	Senescence; size; seed protein content	Dev and morph; seed biochemistry	AT-hook	Early senescence; reduced size; altered seed protein content	604	(87-100)
605	G2383	Senescence; seed protein content	Dev and morph; seed biochemistry	TEO	Early senescence; altered seed protein content	606	(89-149)
607	G571	Senescence; flowering time	Dev and morph; flowering time	bZIP	Delayed senescence; late flowering	608	(160-220)

Table 4

609	G636	Senescence; size	Dev and morph	TH	Premature senescence; reduced size	610	(55-145, 405-498)
611	G878	Senescence; flowering time	Dev and morph; flowering time	WRKY	Delayed senescence; late flowering	612	(250-305, 415-475)
613	G1134	Siliques	Dev and morph	HLH/MYC	Siliques with altered shape	614	(198-247)
615	G1008	Size	Dev and morph	AP2	Small plant	616	(96-163)
617	G1020	Size	Dev and morph	AP2	Very small T1 plants	618	(28-95)
619	G1023	Size	Dev and morph	AP2	Reduced size	620	(128-195)
621	G1053	Size	Dev and morph	bZIP	Small plant	622	(74-120)
623	G1137	Size	Dev and morph	HLH/MYC	Small T1 plants	624	(264-314)
625	G1181	Size	Dev and morph	HS	Small T1 plants	626	(24-114)
627	G1228	Size	Dev and morph	HLH/MYC	Reduced size	628	(179-233)
629	G1277	Size	Dev and morph	AP2	Small plant	630	(18-85)
631	G1309	Size	Dev and morph	MYB-(R1)R2R3	Small plant	632	(9-114)
633	G1314	Size; sugar sensing; seed protein content	Dev and morph; sensing; seed biochemistry	MYB-(R1)R2R3	Reduced size; reduced seedling vigor on high glucose; altered seed protein content	634	(14-116)
635	G1317	Size	Dev and morph	MYB-(R1)R2R3	Reduced size	636	(13-118)
637	G1323	Size; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Small T1 plants, dark green; decreased seed oil, increased seed protein	638	(15-116)
639	G1332	Size; trichome; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Reduced size; reduced trichome density; altered seed oil; and protein content	640	(13-116)
641	G1334	Size	Dev and morph	CAAT	Small, dark green	642	(18-190)
643	G1381	Size	Dev and morph	AP2	Reduced size	644	(68-135)
645	G1382	Size	Dev and morph	WRKY	Small plant	646	(210-266, 385-437)
647	G1435	Size; flowering time	Dev and morph; flowering time	GARP	Increased plant size; late flowering	648	(146-194)
649	G1537	Size	Dev and morph	HB	Small T1 plants with altered development	650	(14-74)
651	G1545	Size	Dev and morph	HB	Reduced size	652	(54-117)
653	G1641	Size; seed oil and protein content	Dev and morph; seed biochemistry	MYB-related	Small plant; altered seed oil and protein content	654	(139-200)
655	G165	Size; seed protein content	Dev and morph; seed biochemistry	MADS	Reduced size; altered seed protein content	656	(7-62)

Table 4

657	G1652	Size; seed oil and protein content	Dev and morph; seed biochemistry	HLH/MYC	Reduced size; altered seed oil and protein content	658	(143-215)
659	G1655	Size	Dev and morph	HLH/MYC	Small plant	660	(134-192)
661	G1671	Size	Dev and morph	NAC	Reduced size	662	(TBD)
663	G1756	Size; seed protein content	Dev and morph; seed biochemistry	WRKY	Reduced size; altered seed protein content	664	(TBD)
665	G1757	Size; seed protein content	Dev and morph; seed biochemistry	WRKY	Small plant; altered seed protein content	666	(158-218)
667	G1782	Size	Dev and morph	CAAT	Small, spindly plant	668	(166-238)
669	G184	Size	Dev and morph	WRKY	Small plant	670	(295-352)
671	G1845	Size	Dev and morph	AP2	Small plant	672	(140-207)
673	G1879	Size; seed oil and protein content	Dev and morph; seed biochemistry	HLH/MYC	Reduced size; altered seed oil and protein content	674	(107-176)
675	G1888	Size	Dev and morph	Z-CO-like	Reduced size, dark green leaves	676	(5-50)
677	G189	Size; seed protein content	Dev and morph; seed biochemistry	WRKY	Increased leaf size; altered seed protein content	678	(240-297)
679	G1939	Size	Dev and morph	PCF	Reduced size	680	(40-102)
681	G194	Size	Dev and morph	WRKY	Small plant	682	(174-230)
683	G1943	Size	Dev and morph	HLH/MYC	Reduced size	684	(335-406)
685	G21	Size; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Reduced size; altered seed oil and protein content	686	(97-164)
687	G2132	Size; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Reduced size; altered seed oil and protein content	688	(TBD)
689	G2145	Size	Dev and morph	HLH/MYC	Reduced size	690	(166-243)
691	G23	Size	Dev and morph	AP2	Small T1 plants	692	(61-117)
693	G2313	Size	Dev and morph	MYB-related	Reduced size	694	(TBD)
695	G2344	Size	Dev and morph	CAAT	Reduced size, slow growth	696	(TBD)
697	G2430	Size	Dev and morph	GARP	Increased leaf size, faster development	698	(425-478)
699	G2517	Size	Dev and morph	WRKY	Reduced size	700	(118-234)
701	G2521	Size	Dev and morph	HLH/MYC	Reduced size	702	(145-213)
703	G258	Size; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Reduced size; altered seed oil and protein content	704	(24-124)

Table 4

705	G280	Size; seed protein content	Dev and morph; seed biochemistry	AT-hook	Reduced size; altered seed protein content	706	(97-104, 130-137-155-162, 185-192)
707	G3	Size	Dev and morph	AP2	Small plant	708	(28-95)
709	G343	Size	Dev and morph	GATA/Zn	Small plant	710	(178-214)
711	G363	Size	Dev and morph	Z-C2H2	Small plant	712	(87-108)
713	G370	Size	Dev and morph	Z-C2H2	Reduced size, shiny leaves	714	(97-117)
715	G385	Size	Dev and morph	HB	Small plant, short inflorescence stems, dark green	716	(60-123)
717	G439	Size	Dev and morph	AP2	Small plant	718	(110-177)
719	G440	Size	Dev and morph	AP2	Small plant	720	(122-189)
721	G5	Size	Dev and morph	AP2	Small plant	722	(149-216)
723	G550	Size	Dev and morph	Z-Dof	Small plant	724	(134-180)
725	G670	Size	Dev and morph	MYB-(R1)R2R3	Small plant	726	(14-122)
727	G760	Size	Dev and morph	NAC	Reduced size	728	(12-156)
729	G831	Size	Dev and morph	AKR	Reduced size	730	(470-591)
731	G864	Size	Dev and morph	AP2	Small plant	732	(119-186)
733	G884	Size	Dev and morph	WRKY	Reduced size	734	(227-285, 407-465)
735	G898	Size; seed oil and protein content	Dev and morph; seed biochemistry	RING/C3HC4	Reduced size; altered seed oil and protein content	736	(148-185)
737	G900	Size	Dev and morph	Z-CO-like	Reduced size	738	(6-28, 48-74)
739	G913	Size; flowering time	Dev and morph; flowering time	AP2	Small plant; late flowering	740	(62-128)
741	G937	Size	Dev and morph	GARP	Slightly reduced size	742	(197-246)
743	G960	Size	Dev and morph	NAC	Small plant	744	(13-156)
745	G991	Size; seed oil and protein content	Dev and morph; seed biochemistry	IAA	Slightly reduced size; altered seed oil and protein content	746	(7-14, 48-59, 82-115, 128-164)
747	G748	Stem; flowering time	Dev and morph; flowering time	Z-Dof	More vascular bundles in stem; late flowering	748	(112-140)
749	G247	Trichome; seed protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Altered trichome distribution; altered seed protein content	750	(15-116)
751	G585	Trichome	Dev and morph	HLH/MYC	Reduced trichome density	752	(436-501)
753	G634	Trichome; seed protein content	Dev and morph; seed biochemistry	TH	Increased trichome density and size; altered seed protein content	754	(62-147, 189-245)
755	G676	Trichome	Dev and morph	MYB-(R1)R2R3	Reduced trichomes	756	(17-119)

Table 4

757	G682	Trichome	Dev and morph	MYB-related	Glabrous, lack of trichomes	758	(27-63)
759	G635	Variegation	Dev and morph	TH		760	(239-323)
761	G1068	Sugar sensing; seed oil and seed	Sugar sensing; seed	AT-hook	Reduced cotyledon expansion in glucose	762	(143-150)
763	G1225	protein content	biochemistry	HLH/MYC	Better germination on sucrose and glucose media; altered seed oil and protein content	764	(78-147)
765	G1337	Sugar sensing	Sugar sensing	Z-CO-like	Decreased germination on sucrose medium	766	(9-75)
767	G1759	Sugar sensing	Sugar sensing	MADS	Reduced germination on high glucose	768	(2-57)
769	G1804	flowering time	flowering time	bZIP	Altered sugar sensing; late flowering	770	(357-407)
771	G207	Sugar sensing	Sugar sensing	MYB-(R1)R2R3	Decreased germination on glucose medium	772	(6-106)
773	G218	Sugar sensing; seed oil content	Sugar sensing; seed	MYB-(R1)R2R3	Reduced cotyledon expansion in glucose; altered seed oil content	774	(TBD)
775	G241	Sugar sensing; seed oil and seed	Sugar sensing; seed		Decreased germination and growth on glucose medium; decreased seed oil, altered protein content	776	(14-114)
777	G254	protein content	biochemistry	MYB-(R1)R2R3	Decreased germination and growth on glucose medium	778	(62-106)
779	G26	Sugar sensing	Sugar sensing	MYB-related AP2	Decreased germination and growth on glucose medium	780	(67-134)
781	G263	Sugar sensing	Sugar sensing		Decreased root growth on sucrose medium, root specific expression	782	(TBD)
783	G308	Sugar sensing	Sugar sensing	HS	No germination on glucose medium	784	(270-274)
785	G38	Sugar sensing	Sugar sensing	SCR	Reduced germination on glucose medium	786	(76-143)
787	G43	Sugar sensing	Sugar sensing	AP2	Decreased germination and growth on glucose medium	788	(104-172)
789	G536	Sugar sensing	Sugar sensing	GF14	Decreased germination and growth on glucose medium	790	(226-233)
791	G567	Sugar sensing; seed oil and seed	Sugar sensing; seed		Decreased seedling vigor on high glucose; altered seed oil and protein content	792	(210-270)
793	G680	protein content	biochemistry	bZIP		794	(24-70)
795	G867	flowering time	flowering time	MYB-related	Reduced germination on glucose medium; late flowering	796	(59-124)
797	G956	Sugar sensing	Sugar sensing	AP2	Better seedling vigor on sucrose medium	798	(TBD)
799	G996	Sugar sensing	Sugar sensing	NAC	Reduced germination on glucose medium	800	(14-114)
801	G1946	Seed glucosinolates, oil, protein content	Seed biochemistry	MYB-(R1)R2R3	Increased germination on glucose medium		
803	G217	Seed oil composition	Seed biochemistry	HS	Increase in M3950; increased oil content; decreased protein content	802	(32-130)
				MYB-related	Increase in 20:2	804	(8-67)

Table 4

805	G2192	Seed oil composition	Seed biochemistry	bZIP-NIN	Altered composition	806	(600-700)
807	G504	Seed oil composition;	Seed biochemistry	NAC	Altered seed oil composition and content; altered seed protein content	808	(TBD)
809	G622	Seed oil composition	Seed biochemistry	ABI3/NP-1	Decreased 18:2 fatty acid	810	(TBD)
811	G778	Seed oil composition	Seed biochemistry	HLH/MYC	Increased seed 18:1 fatty acid	812	(220-267)
813	G791	Seed oil composition	Seed biochemistry	HLH/MYC	Altered seed fatty acid composition	814	(75-143)
815	G861	Seed oil composition; seed oil content	Seed biochemistry	MADS	Increase in 16:1; altered seed oil content	816	(2-57)
817	G938	Seed oil composition	Seed biochemistry	EIL	Altered seed fatty acid composition	818	(96-104)
819	G965	Seed oil composition	Seed biochemistry	HB	Increase in 18:1	820	(423-486)
821	G1143	Seed oil and protein content	Seed biochemistry	HLH/MYC	Altered seed oil and protein content	822	(33-82)
823	G1190	Seed oil content	Seed biochemistry	AKR	Increased content	824	(entire protein)
825	G1198	Seed oil and protein content	Seed biochemistry	bZIP	Altered seed oil and protein content	826	(173-223)
827	G1226	Seed oil and protein content	Seed biochemistry	HLH/MYC	Altered seed oil and protein content	828	(115-174)
829	G1451	Seed oil content	Seed biochemistry	ARF	Altered seed oil content	830	(22-357)
831	G1478	Seed oil and protein content; flowering time	Seed biochemistry; flowering time	Z-CO-like	Altered seed oil, protein content; late flowering	832	(32-76)
833	G1496	Seed oil content	Seed biochemistry	HLH/MYC	Altered seed oil content	834	(184-248)
835	G1526	Seed oil content	Seed biochemistry	SWI/SNF	Increased seed oil content	836	(493-620, 864-1006)
837	G1543	Seed oil content	Seed biochemistry	HB	Decreased seed oil	838	(135-195)
839	G162	Seed oil and protein content	Seed biochemistry	MADS	Altered seed oil content; altered seed oil and protein content	840	(2-57)

Table 4

841	G1640	Seed oil content; Seed oil and protein content	Seed biochemistry	MYB-(R1)R2R3	Increased seed oil	842	(14-115)
843	G1644	Seed oil and protein content	Seed biochemistry	MYB-(R1)R2R3	Altered seed oil, protein content	844	(39-102)
845	G1646	Seed oil content; Seed oil and protein content	Seed biochemistry	CAAT	Altered seed oil content	846	(72-162)
847	G1672	Seed oil content; Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil content	848	(41-194)
849	G1677	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil, protein content	850	(17-181)
851	G1765	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	852	(20-140)
853	G1777	Seed oil and protein content	Seed biochemistry	RING/C3HC4	Increased oil, decreased protein content	854	(124-247)
855	G1793	Seed oil content; Seed oil and protein content	Seed biochemistry	AP2	Increased seed oil content	856	(179-255, 281-349)
857	G180	Seed oil content; Seed oil and protein content; flowering time	Seed biochemistry	WRKY	Decreased seed oil content	858	(118-174)
859	G192	Seed oil and protein content; flowering time	Seed biochemistry; flowering time	WRKY	Altered seed oil and protein content; late flowering	860	(128-185)
861	G1948	Seed oil and protein content	Seed biochemistry	AKR	Altered seed oil and protein content	862	(entire protein)
863	G2123	Seed oil and protein content	Seed biochemistry	GF14	Altered seed oil and protein content	864	(99-109)
865	G2138	Seed oil content; Seed oil and protein content	Seed biochemistry	AP2	Increased seed oil content	866	(TBD)
867	G2139	Seed oil content; Seed oil and protein content	Seed biochemistry	MADS	Increased seed content	868	(14-69)
869	G2343	Seed oil content; Seed oil and protein content	Seed biochemistry	MYB-(R1)R2R3	Altered seed oil content	870	(14-116)
871	G265	Seed oil and protein content	Seed biochemistry	HS	Altered seed oil and protein content	872	(11-105)
873	G2792	Seed oil content; Seed oil and protein content	Seed biochemistry	HLH/MYC	Increased seed oil content	874	(190-258)
875	G2830	Seed oil and protein content	Seed biochemistry	Z-C2H2	Altered seed oil and protein content	876	(245-266)
877	G286	Seed oil and protein content	Seed biochemistry	ENBP	Altered seed oil and protein content	878	(TBD)

Table 4

879	G291	Seed oil content	Seed biochemistry	MISC	Increased seed oil content	880	(132-160)
881	G427	Seed oil and protein content	Seed biochemistry	HB	Increased oil content; decreased protein content	882	(307-370)
883	G509	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	884	(13-169)
885	G519	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	886	(11-104)
887	G561	Seed oil content	Seed biochemistry	bZIP	Altered seed oil content	888	(248-308)
889	G590	Seed oil and protein content	Seed biochemistry	HLH/MYC	Altered seed oil and protein content	890	(202-254)
891	G818	Seed oil content	Seed biochemistry	HS	Increased content	892	(70-162)
893	G849	Seed oil and protein content	Seed biochemistry	BPF-1	Increased seed oil, altered protein content	894	(324-413, 504-583)
895	G892	Seed oil and protein content	Seed biochemistry	RING/C3H2C3	Altered seed oil, protein content	896	(177-270)
897	G961	Seed oil content	Seed biochemistry	NAC	Altered seed oil content	898	(15-140)
899	G1465	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	900	(242-306)
901	G425	Seed oil content	Seed biochemistry	HB	Altered seed oil content	902	(TBD)
903	G347	Seed oil and protein content	Seed biochemistry	Z-LSDlike	Altered seed oil and protein content	904	(9-39, 50-70, 80-127)
905	G1512	Seed oil and protein content	Seed biochemistry	RING/C3HC4	Altered seed oil and protein content	906	(39-93)
907	G2069	Seed oil and protein content	Seed biochemistry	bZIP	Altered seed oil and protein content	908	(TBD)
909	G1852	Seed oil content	Seed biochemistry	AKR	Altered seed oil content	910	(1-801)
911	G1793	Seed oil content	Seed biochemistry	AP2	Altered seed oil content	912	(179-255, 281-349)
913	G761	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	914	(10-156)
915	G1056	Seed oil content	Seed biochemistry	bZIP	Altered seed oil content	916	(183-246)
917	G1447	Seed oil content	Seed biochemistry	MISC	Altered seed oil content	918	(3-54, 124-156)

Table 4

919	G323	Seed oil and protein content	Seed biochemistry	RING/C3HC4	Altered seed oil and protein content	920	(48-96)
921	G176	Seed oil content	Seed biochemistry	WRKY	Altered seed oil content	922	(117-173, 234-290)
923	G174	Seed oil and protein content	Seed biochemistry	WRKY	Altered seed oil and protein content	924	(111-166, 283-339)
925	G715	Seed oil content	Seed biochemistry	CAAT	Altered seed oil content	926	(60-132)
927	G588	Seed oil and protein content	Seed biochemistry	HLH/MYC	Altered seed oil and protein content	928	(309-376)
929	G1758	Seed oil and protein content	Seed biochemistry	WRKY	Altered seed oil and protein content	930	(109-165)
931	G2148	Seed oil content	Seed biochemistry	HLH/MYC	Altered seed oil content	932	(130-268)
933	G2379	Seed oil content	Seed biochemistry	TH	Altered seed oil content	934	(19-110, 173-232)
935	G1462	Seed oil content	Seed biochemistry	NAC	Altered seed oil content	936	(TBD)
937	G1211	Seed oil and protein content	Seed biochemistry	MISC	Altered seed oil and protein content	938	(123-179)
939	G1048	Seed oil content	Seed biochemistry	bZIP	Altered seed oil content	940	(138-190)
941	G986	Seed oil content	Seed biochemistry	WRKY	Altered seed oil content	942	(146-203)
943	G789	Seed oil content	Seed biochemistry	HLH/MYC	Altered seed oil content	944	(253-313)
945	G2085	Seed oil and protein content	Seed biochemistry	RING/C3HC4	Altered seed oil and protein content	946	(TBD)
947	G1783	Seed oil and protein content	Seed biochemistry	MYB-related	Altered seed oil and protein content	948	(81-129)
949	G2072	Seed oil and protein content	Seed biochemistry	bZIP	Altered seed oil and protein content	950	(90-149)
951	G931	Seed oil and protein content	Seed biochemistry	CAAT	Altered seed oil and protein content	952	(TBD)
953	G278	Seed oil and protein content	Seed biochemistry	AKR	Altered seed oil and protein content	954	(2-593)
955	G2421	Seed oil content	Seed biochemistry	MYB-(R1)R2R3	Altered seed oil content	956	(9-110)
957	G2032	Seed oil content	Seed biochemistry	AKR	Altered seed oil content	958	(entire protein)

Table 4

959	G1396	Seed oil and protein content	Seed biochemistry	SIFA	Altered seed oil and protein content	960	(TBD)
961	G819	Seed oil and protein content	Seed biochemistry	ARF	Altered seed oil and protein content	962	(64-406)
963	G2295	Seed oil content	Seed biochemistry	MADS	Altered seed oil content	964	(2-57)
965	G312	Seed oil content	Seed biochemistry	SCR	Altered seed oil content	966	(320-336)
967	G1444	Seed oil and protein content	Seed biochemistry	GRF-like	Altered seed oil and protein content	968	(168-193)
969	G801	Seed oil content	Seed biochemistry	PCF	Altered seed oil content	970	(32-93)
971	G1950	Seed oil content	Seed biochemistry	AKR	Altered seed oil content	972	(65-228)
973	G958	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	974	(7-156)
975	G1037	Seed oil and protein content	Seed biochemistry	GARP	Altered seed oil and protein content	976	(11-134, 200-248)
977	G2065	Seed oil content	Seed biochemistry	MADS	Altered seed oil content	978	(TBD)
979	G2137	Seed oil and protein content	Seed biochemistry	WRKY	Altered seed oil and protein content	980	(109-168)
981	G746	Seed oil content	Seed biochemistry	RING/C3HC4	Altered seed oil content	982	(139-178)
983	G2701	Seed oil and protein content	Seed biochemistry	MYB-related	Altered seed oil and protein content	984	(33-81, 129-183)
985	G1819	Seed oil content	Seed biochemistry	CAAT	Altered seed oil content	986	(46-188)
987	G1227	Seed oil and protein content	Seed biochemistry	HLL/MYC	Altered seed oil and protein content	988	(183-244)
989	G2417	Seed oil content	Seed biochemistry	GARP	Altered seed oil content	990	(235-285)
991	G2116	Seed oil content	Seed biochemistry	bZIP	Altered seed oil content	992	(150-210)
993	G647	Seed oil content	Seed biochemistry	Z-C3H	Altered seed oil content	994	(77-192)
995	G974	Seed oil and protein content	Seed biochemistry	AP2	Altered seed oil and protein content	996	(81-140)
997	G1419	Seed protein content	Seed biochemistry	AP2	Increased seed protein	998	(69-137)

Table 4

999	G1634	Seed protein content	Seed biochemistry	MYB-related	Altered seed protein content	1000	(129-180)
1001	G1637	Seed protein content	Seed biochemistry	MYB-related	Altered seed protein content	1002	(109-173)
1003	G1818	Seed protein content; flowering time	Seed biochemistry; flowering time	CAAT	Increased protein content; late flowering	1004	(36-113)
1005	G1820	Seed oil and protein content	Seed biochemistry	CAAT	Altered seed oil, protein content	1006	(70-133)
1007	G1903	Seed oil and protein content	Seed biochemistry	Z-Dof	Altered seed oil and protein content	1008	(134-180)
1009	G371	Seed oil and protein content	Seed biochemistry	RING/C3HC4	Altered seed oil and protein content	1010	(21-74)
1011	G597	Seed protein content	Seed biochemistry	AT-hook	Altered seed protein content	1012	(97-104, 137-144)
1013	G1009	Seed protein content	Seed biochemistry	AP2	Altered seed protein content	1014	(201-277, 303-371)
1015	G170	Seed protein content	Seed biochemistry	MADS	Altered seed protein content	1016	(2-57)
1017	G1768	Seed protein content	Seed biochemistry	SCR	Altered seed protein content	1018	(54-413)
1019	G185	Seed protein content	Seed biochemistry	WRKY	Altered seed protein content	1020	(113-172)
1021	G1931	Seed protein content	Seed biochemistry	WRKY	Altered seed protein content	1022	(114-170)
1023	G2543	Seed protein content	Seed biochemistry	HB	Altered seed protein content	1024	(31-91)
1025	G264	Seed protein content	Seed biochemistry	HS	Altered seed protein content	1026	(24-114)
1027	G32	Seed protein content	Seed biochemistry	AP2	Altered seed protein content	1028	(17-84)
1029	G436	Seed protein content	Seed biochemistry	HB	Altered seed protein content	1030	(22-85)
1031	G556	Seed protein content	Seed biochemistry	bZIP	Altered seed protein content	1032	(83-143)
1033	G1420	Seed protein content	Seed biochemistry	WRKY	Altered seed protein content	1034	(221-280)
1035	G1412	Seed protein content	Seed biochemistry	NAC	Altered seed protein content	1036	(17-159)

Table 4

1037	G738	Seed protein content	Seed biochemistry	Z-Dof	Altered seed protein content	1038	(351-393)
1039	G2426	Seed protein content	Seed biochemistry	MYB-(R1)R2R3	Altered seed protein content	1040	(14-114)
1041	G1524	Seed protein content	Seed biochemistry	RING/C3HC4	Altered seed protein content	1042	(49-110)
1043	G1243	Seed protein content	Seed biochemistry	SWI/SNF	Altered seed protein content	1044	(216-609)
1045	G631	Seed protein content	Seed biochemistry	bZIP	Altered seed protein content	1046	(TBD)
1047	G1909	Seed protein content	Seed biochemistry	Z-Dof	Altered seed protein content	1048	(23-51)
1049	G1663	Seed protein content	Seed biochemistry	PCF	Altered seed protein content	1050	(TBD)
1051	G1231	Seed protein content	Seed biochemistry	Z-C4HC3	Altered seed protein content	1052	(TBD)
1053	G227	Seed protein content	Seed biochemistry	MYB-(R1)R2R3	Altered seed protein content	1054	(13-112)
1055	G1842	Seed protein content	Seed biochemistry	MADS	Altered seed protein content	1056	(2-57)
1057	G1505	Seed protein content	Seed biochemistry	GATA/Zn	Altered seed protein content	1058	(TBD)
1059	G657	Seed protein content	Seed biochemistry	MYB-(R1)R2R3	Altered seed protein content	1060	(TBD)
1061	G1959	Seed protein content	Seed biochemistry	GARP	Altered seed protein content	1062	(46-97)
1063	G2180	Seed protein content	Seed biochemistry	NAC	Altered seed protein content	1064	(7-156)
1065	G1817	Seed protein content	Seed biochemistry	PMR	Altered seed protein content	1066	(47-331)
1067	G1649	Seed protein content	Seed biochemistry	HLH/MYC	Altered seed protein content	1068	(225-295)
1069	G2131	Seed protein content	Seed biochemistry	AP2	Altered seed protein content	1070	(50-186, 112-183)
1071	G215	Seed protein content	Seed biochemistry	MYB-related	Altered seed protein content	1072	(TBD)
1073	G1508	Seed protein content	Seed biochemistry	GATA/Zn	Altered seed protein content	1074	(38-63)
1075	G2110	Seed protein content	Seed biochemistry	WRKY	Altered seed protein content	1076	(239-298)

Table 4

1077	G2442	Seed protein content	Seed biochemistry	RING/C3HC4	Altered seed protein content	1078	(220-246)
1079	G1051	Flowering time	Flowering time	bZIP	Late flowering	1080	(189-250)
1081	G1052	Flowering time	Flowering time	bZIP	Late flowering	1082	(201-261)
1083	G1079	Flowering time	Flowering time	BZIP12	Late flowering; altered seed protein content	1084	(1-50)
1085	G1335	Flowering time	Flowering time	Z-CLDSH	Late flowering, slow growth	1086	(24-43, 131-144, 185-203)
1087	G157	Flowering time	Flowering time	MADS	Altered flowering; significant overexpression delays	1088	(2-57)
1089	G1895	Flowering time	Flowering time	Z-Dof	Late flowering	1090	(55-110)
1091	G1900	Flowering time	Flowering time	Z-Dof	Late flowering	1092	(54-106)
1093	G2007	Flowering time; seed protein content	Flowering time; seed				
1095	G214	Flowering time	Flowering time	MYB-(R1)R2R3	Late flowering; altered seed protein content	1094	(TBD)
1097	G2155	Flowering time	Flowering time	MYB-related	Late flowering	1096	(22-71)
1099	G234	Flowering time	Flowering time	AT-hook	Late flowering	1098	(18-38)
1101	G361	Flowering time	Flowering time	MYB-(R1)R2R3	Late flowering, small plant	1100	(14-115)
1103	G562	Flowering time	Flowering time	Z-C2H2	Late flowering	1102	(43-63)
1105	G591	Flowering time	Flowering time	bZIP	Late flowering	1104	(253-315)
1107	G8	Flowering time	Flowering time	HLH/MYC	Late flowering	1106	(143-240)
1109	G859	Flowering time; seed protein content	Flowering time; seed	AP2	Late flowering	1108	(151-217, 243-296)
1111	G878	Flowering time	Flowering time	MADS	Late flowering; altered seed protein content	1110	(TBD)
1113	G971	Flowering time	Flowering time	WRKY	Late flowering	1112	(250-305, 415-475)
1115	G975	Flowering time; morphology: other	Flowering time; dev and morph	AP2	Late flowering; glossy leaves	1116	(4-71)
1117	G994	Flowering time	Flowering time	MYB-(R1)R2R3	Late flowering, small	1118	(14-123)
1119	G2347	Flowering time	Flowering time	SBP	Late flowering, small	1120	(60-136)
1121	G2010	Flowering time	Flowering time	SBP	Late flowering	1122	(53-127)

Table 5

SEQ ID NO	GID	Test Sequence ID	Smallest Sum Probability	Test Sequence Species	Test Sequence GenBank Annotation
859	G192	AW596933	7.70E-40	[Glycine max]	sl84f07.y1 Gm-c1034 Glycine max cDNA clone GENO
859	G192	AV423663	2.40E-39	[Lotus japonicus]	AV423663 Lotus japonicus young plants (two-
859	G192	BI422074	4.50E-34	[Lycopersicon esculentum]	EST532740 tomato callus, TAMU Lycop
859	G192	AW447931	1.40E-27	[Triticum aestivum]	BRY_1082 BRY Triticum aestivum cDNA clone
859	G192	BE98060	2.60E-24	[Medicago truncatula]	EST429783 GVSN Medicago truncatula cDNA
859	G192	AC018727	1.70E-23	[Oryza sativa]	chromosome 10 clone OSJNBa0056G17, *** SEQUENC
859	G192	BG600477	1.00E-20	[Solanum tuberosum]	EST505372 cSTS Solanum tuberosum cDNA clo
859	G192	BG356878	2.80E-16	[Sorghum bicolor]	OV2_11_B04.g1_A002 Ovary 2 (OV2) Sorghum bi
859	G192	gi12039364	1.10E-31	[Oryza sativa]	putative DNA-binding protein.
859	G192	gi4894963	3.30E-14	[Avena sativa]	DNA-binding protein WRKY3.
859	G192	gi1432056	5.80E-14	[Petroselinum crispum]	WRKY3.
859	G192	gi4760596	2.60E-13	[Nicotiana tabacum]	DNA-binding protein NWRKY3.
859	G192	gi11993901	1.40E-12	[Dactylis glomerata]	somatic embryogenesis related protein.
859	G192	gi927025	7.60E-09	[Cucumis sativus]	SPF1-like DNA-binding protein.
859	G192	gi13620227	8.40E-09	[Lycopersicon esculentum]	hypothetical protein.
859	G192	gi3420906	2.80E-08	[Pimpinella brachycarpa]	zinc finger protein; WRKY1.
859	G192	gi1159877	4.70E-08	[Avena fatua]	DNA-binding protein.
859	G192	gi484261	1.60E-07	[Ipomoea batatas]	SPF1 protein.
801	G1946	LPHSF8	1.10E-119	[Lycopersicon peruvianum]	L.peruvianum Lp-hsf8 mRNA for heat
801	G1946	AC087771	4.10E-112	[Medicago truncatula]	clone 8D15, *** SEQUENCING IN PROGRESS
801	G1946	LEHSF8	5.90E-103	[Lycopersicon esculentum]	L.esculentum Le-hsf8 gene for heat
801	G1946	AW569138	3.10E-75	[Glycine max]	sl63g09.y1 Gm-r1030 Glycine max cDNA clone GENO
801	G1946	BG890899	1.30E-70	[Solanum tuberosum]	EST516750 cSTD Solanum tuberosum cDNA clo
801	G1946	AC027658	4.60E-53	[Oryza sativa]	subsp. japonica BAC nbx0006113, chromosome 10
801	G1946	AV833112	4.90E-52	[Hordeum vulgare subsp. vulgare]	AV833112 K. Sato unpublished
801	G1946	gi19492	2.80E-121	[Lycopersicon peruvianum]	heat shock transcription factor 8
801	G1946	gi19260	5.10E-106	[Lycopersicon esculentum]	heat stress transcription factor
801	G1946	gi662924	2.00E-47	[Glycine max]	heat shock transcription factor 21.
801	G1946	gi5821138	9.70E-46	[Nicotiana tabacum]	heat shock factor.
801	G1946	gi11761077	2.90E-40	[Oryza sativa]	putative heat shock factor protein 1 (HSF 1)
801	G1946	gi886742	3.20E-40	[Zea mays]	heat shock factor.
801	G1946	gi7158882	2.70E-38	[Medicago sativa]	heat shock transcription factor.
801	G1946	gi3550588	1.90E-30	[Pisum sativum]	heat shock transcription factor (HSFA).

Table 5

801	G1946	gi100546	0.46	[Avena sativa]	avenin precursor - oat.
801	G1946	gi14190783	1	[Apium graveolens]	putative phloem transcription factor M1.
239	G375	AW696439	3.40E-33	[Medicago truncatula]	NF106B07ST1F1060 Developing stem Medica
239	G375	BG595870	1.90E-31	[Solanum tuberosum]	EST494548 cSTS Solanum tuberosum cDNA clo
239	G375	AI899263	3.70E-31	[Lycopersicon esculentum]	EST288706 tomato ovary, TAMU Lycope
239	G375	NTBBF3	4.00E-31	[Nicotiana tabacum]	N.tabacum mRNA for zinc finger protein, B
239	G375	BG405482	2.70E-30	[Glycine max]	sac44a11.y1 Gm-c1062 Glycine max cDNA clone GEN
239	G375	AB028130	3.30E-30	[Oryza sativa]	mRNA for Dof zinc finger protein, complete cds
239	G375	AB026297	7.30E-28	[Pisum sativum]	mRNA for elicitor-responsive Dof protein ERDP
239	G375	HVBPBF	1.10E-27	[Hordeum vulgare]	mRNA for DNA binding protein BPBF.
239	G375	BG263089	1.70E-27	[Triticum aestivum]	WHE2337_A02_A03ZS Wheat pre-anthesis spik
239	G375	ZMU82230	4.20E-27	[Zea mays]	endosperm-specific prolamin box binding factor (PB
239	G375	gi4996640	1.90E-37	[Oryza sativa]	Dof zinc finger protein.
239	G375	gi3777436	8.10E-35	[Hordeum vulgare]	DNA binding protein.
239	G375	gi2393775	1.10E-33	[Zea mays]	prolamin box binding factor.
239	G375	gi1360088	2.00E-33	[Nicotiana tabacum]	Zn finger protein.
239	G375	gi3790264	4.30E-32	[Triticum aestivum]	PBF protein.
239	G375	gi6092016	1.30E-29	[Pisum sativum]	elicitor-responsive Dof protein ERDP.
239	G375	gi7688355	5.60E-29	[Solanum tuberosum]	Dof zinc finger protein.
239	G375	gi1669341	4.60E-20	[Cucurbita maxima]	AOBP (ascorbate oxidase promoter-binding
239	G375	gi3929325	5.50E-18	[Dendrobium grex Madame Thong-In]	putative DNA-binding prot
239	G375	gi19547	5.50E-06	[Medicago sativa subsp. falcata]	environmental stress and a
273	G1255	AC087181	1.60E-46	[Oryza sativa]	chromosome 3 clone OSJNBa0018H01, *** SEQUENCI
273	G1255	BG239774	4.50E-33	[Glycine max]	sab74c03.y1 Gm-c1032 Glycine max cDNA clone GEN
273	G1255	BG321336	1.70E-32	[Descurainia sophia]	Ds01_06h10_A Ds01_AAFc ECORC cold stress
273	G1255	AI772841	2.90E-30	[Lycopersicon esculentum]	EST253941 tomato resistant, Cornell
273	G1255	BF480245	4.60E-29	[Mesembryanthemum crystallinum]	L0-2152T3 Ice plant Lambda Un
273	G1255	AW688119	2.10E-28	[Medicago truncatula]	NF002E07ST1F1000 Developing stem Medica
273	G1255	BF266327	1.80E-26	[Hordeum vulgare]	HV_CEA0014N02f Hordeum vulgare seedling gre
273	G1255	AW671538	5.80E-25	[Sorghum bicolor]	LG1_348_B08.b1_A002 Light Grown 1 (LG1) Sor
273	G1255	BI072021	5.30E-20	[Populus tremula x Populus tremuloides]	C067P76U Populus stra
273	G1255	BG273908	4.90E-19	[Vitis vinifera]	EST 110 Green Grape berries Lambda Zap II Li
273	G1255	gi13702811	3.70E-52	[Oryza sativa]	putative zinc finger protein.
273	G1255	gi11037311	4.00E-21	[Brassica nigra]	constans-like protein.
273	G1255	gi2303683	1.10E-19	[Brassica napus]	unnamed protein product.
273	G1255	gi4091804	2.30E-18	[Malus x domestica]	CONSTANS-like protein 1.

Table 5

273	G1255	gi3341723	4.30E-18	[Raphanus sativus]	CONSTANS-like 1 protein.
273	G1255	gi10946337	5.20E-17	[Ipomoea nil]	CONSTANS-like protein.
273	G1255	gi4557093	3.30E-15	[Pinus radiata]	zinc finger protein.
273	G1255	gi8132543	0.97	[Chloroplast Zamia furfuracea]	cytochrome b559 alpha subunit
273	G1255	gi11795	0.99	[Nicotiana tabacum]	put. psbE protein (aa 1-83).
273	G1255	gi65646	0.99	[Chloroplast Nicotiana tabacum]	cytochrome b559 component p
557	G865	BE419451	3.70E-32	[Triticum aestivum]	WWS012.C2R000101 ITEC WWS Wheat Scutellum
557	G865	AW560968	1.10E-28	[Medicago truncatula]	EST316016 DSIR Medicago truncatula cDNA
557	G865	AW782252	1.20E-26	[Glycine max]	sm03d11.y1 Gm-c1027 Glycine max cDNA clone GENO
557	G865	BI421895	3.60E-25	[Lycopersicon esculentum]	EST532561 tomato callus, TAMU Lycop
557	G865	BE642320	1.60E-24	[Ceratopteris richardii]	Cri2_5_L17_SP6 Ceratopteris Spore Li
557	G865	BE494041	1.60E-24	[Secale cereale]	WHE1277_B09_D17ZS Secale cereale anther cDNA
557	G865	D39914	2.60E-24	[Oryza sativa]	RICS1576A Rice shoot Oryza sativa cDNA, mRNA s
557	G865	AV428124	9.00E-23	[Lotus japonicus]	AV428124 Lotus japonicus young plants (two- Tobacco mRNA for EREBP-2, complete cds.
557	G865	TOBBY4D	1.80E-21	[Nicotiana tabacum]	ERF1.
557	G865	gi1208495	2.40E-23	[Nicotiana tabacum]	ethylene-responsive element binding
557	G865	gi8809571	5.10E-23	[Nicotiana sylvestris]	Pt14.
557	G865	gi3342211	1.40E-22	[Lycopersicon esculentum]	AP2-related transcription f
557	G865	gi7528276	1.70E-22	[Mesembryanthemum crystallinum]	Putative AP2 domain containing protein.
557	G865	gi15217291	7.80E-22	[Oryza sativa]	AP2 domain containing protein.
557	G865	gi3264767	2.70E-21	[Prunus armeniaca]	AP2-domain DNA-binding protein.
557	G865	gi8980313	2.10E-20	[Catharanthus roseus]	apetala2 domain-containing protein.
557	G865	gi8571476	9.30E-20	[Atriplex hortensis]	DNA binding protein homolog.
557	G865	gi1688233	1.40E-19	[Solanum tuberosum]	ethylene-responsive element binding
557	G865	gi6478845	1.80E-19	[Matricaria chamomilla]	BOHOJ67TR BOHO Brassica oleracea genomic
23	G2509	BH577856	2.50E-29	[Brassica oleracea]	sak01e08.y1 Gm-c1074 Glycine max cDNA clone SOY
23	G2509	BM269574	5.90E-28	[Glycine max]	WWS012.C2R000101 ITEC WWS Wheat Scutellum
23	G2509	BE419451	2.20E-27	[Triticum aestivum]	EST249507 tomato ovary, TAMU Lycop
23	G2509	AI483636	7.80E-27	[Lycopersicon esculentum]	EST316016 DSIR Medicago truncatula cDNA
23	G2509	AW560968	8.90E-27	[Medicago truncatula]	Cri2_5_L17_SP6 Ceratopteris Spore Li
23	G2509	BE642320	4.30E-26	[Ceratopteris richardii]	chromosome 1 clone P0677H08, *** SEQUENCING IN
23	G2509	AP003286	1.00E-25	[Oryza sativa]	WHE1277_B09_D17ZS Secale cereale anther cDNA
23	G2509	BE494041	3.20E-25	[Secale cereale]	HVSMEh0102106f Hordeum vulgare 5-45 DAP spi
23	G2509	BE602106	1.10E-24	[Hordeum vulgare]	AV428124 Lotus japonicus young plants (two- AP2 domain containing protein.
23	G2509	AV428124	1.00E-23	[Lotus japonicus]	
23	G2509	gi3264767	4.00E-27	[Prunus armeniaca]	

Table 5

23	G2509	gi12003376	1.40E-23	[Nicotiana tabacum]	Avr9/Cf-9 rapidly elicited protein 1.
23	G2509	gi14140141	2.30E-23	[Oryza sativa]	putative AP2-related transcription factor.
23	G2509	gi1688233	5.40E-23	[Solanum tuberosum]	DNA binding protein homolog.
23	G2509	gi4099921	2.60E-22	[Stylosanthes hamata]	EREBP-3 homolog.
23	G2509	gi8809571	7.80E-22	[Nicotiana sylvestris]	ethylene-responsive element binding
23	G2509	gi3342211	1.00E-21	[Lycopersicon esculentum]	Pti4.
23	G2509	gi7528276	2.70E-21	[Mesembryanthemum crystallinum]	AP2-related transcription f
23	G2509	gi17385636	1.90E-20	[Matricaria chamomilla]	ethylene-responsive element binding
23	G2509	gi18496063	3.30E-20	[Fagus sylvatica]	ethylene responsive element binding prote
1119	G2347	BI931517	5.30E-31	[Lycopersicon esculentum]	EST551406 tomato flower, 8 mm to pr
1119	G2347	BE058432	4.20E-29	[Glycine max]	sn16a06.y1 Gm-c1016 Glycine max cDNA clone GENO
1119	G2347	AMSPB1	1.80E-28	[Antirrhinum majus]	A.majus mRNA for squamosa-promoter bindin
1119	G2347	BG525285	5.70E-28	[Stevia rebaudiana]	48-3 Stevia field grown leaf cDNA Stevia
1119	G2347	L38193	4.60E-27	[Brassica rapa]	BNAF1025E Mustard flower buds Brassica rapa c
1119	G2347	BG455868	6.40E-27	[Medicago truncatula]	NF068F05PL1F1045 Phosphate starved leaf
1119	G2347	BG097153	1.70E-24	[Solanum tuberosum]	EST461672 potato leaves and petioles Sola
1119	G2347	BF482644	1.60E-23	[Triticum aestivum]	WHE2301-2304 A21 A21ZS Wheat pre-anthesis
1119	G2347	AW747167	2.30E-23	[Sorghum bicolor]	WS1_66 F11.b1 A002 Water-stressed 1 (WS1) S
1119	G2347	BG442540	2.50E-23	[Gossypium arboreum]	GA_Ea0017G06f Gossypium arboreum 7-10 d
1119	G2347	gi1183864	1.50E-31	[Antirrhinum majus]	squamosa-promoter binding protein 2.
1119	G2347	gi5931786	3.40E-25	[Zea mays]	SBP-domain protein 5.
1119	G2347	gi8468036	1.40E-21	[Oryza sativa]	Similar to Arabidopsis thaliana chromosome 2
1119	G2347	gi9087308	6.60E-09	[Mitochondrion Beta vulgaris var. altissima]	orf102a.
1119	G2347	gi7209500	0.83	[Brassica rapa]	S-locus pollen protein.
43	G988	CRU303349	3.10E-208	[Capsella rubella]	ORF1, ORF2, ORF3, ORF4, ORF5 and ORF6 (pa
43	G988	A84072	4.50E-86	[Lycopersicon esculentum]	Sequence 1 from Patent WO9846759.
43	G988	A84080	3.30E-85	[Solanum tuberosum]	Sequence 9 from Patent WO9846759.
43	G988	AP003944	1.30E-57	[Oryza sativa]	chromosome 6 clone OJ1126_F05, *** SEQUENCING
43	G988	AX081276	2.80E-43	[Brassica napus]	Sequence 1 from Patent WO0109356.
43	G988	ZMA242530	1.50E-40	[Zea mays]	partial d8 gene for gibberellin response modula
43	G988	AX005804	2.50E-37	[Triticum aestivum]	Sequence 13 from Patent WO9909174.
43	G988	AB048713	9.10E-33	[Pisum sativum]	PsSCR mRNA for SCARECROW, complete cds.
43	G988	AW774515	2.00E-29	[Medicago truncatula]	EST333666 KV3 Medicago truncatula cDNA
43	G988	BE822458	1.20E-27	[Glycine max]	GM700017A20H12 Gm-1070 Glycine max cDNA clone
43	G988	gi13620166	8.00E-211	[Capsella rubella]	hypothetical protein.
43	G988	gi4160441	1.40E-87	[Lycopersicon esculentum]	lateral suppressor protein.

Table 5

43	G988	gi10178637	2.20E-48	[Zea mays]	SCARECROW.
43	G988	gi6970472	1.20E-47	[Oryza sativa]	OsGAI.
43	G988	gi5640157	2.80E-45	[Triticum aestivum]	gibberellin response modulator.
43	G988	gi13170126	7.10E-45	[Brassica napus]	unnamed protein product.
43	G988	gi13365610	1.10E-40	[Pisum sativum]	SCARECROW.
43	G988	gi14318115	1.10E-14	[Zea mays subsp. mays]	gibberellin response modulator.
43	G988	gi14318165	7.30E-14	[Triticum dactyloides]	gibberellin response modulator.
43	G988	gi347457	2.40E-05	[Glycine max]	hydroxyproline-rich glycoprotein.
459	G2346	AMA011622	3.10E-35	[Antirrhinum majus]	mRNA for squamosa promoter binding
459	G2346	AW691786	1.80E-26	[Medicago truncatula]	NF044B06ST1F1000 Developing stem Medica
459	G2346	AQ273505	7.00E-25	[Oryza sativa]	nbxb0030003f CUGI Rice BAC Library Oryza sativ
459	G2346	AW932595	7.90E-24	[Lycopersicon esculentum]	EST358438 tomato fruit mature green
459	G2346	BG593787	9.50E-24	[Solanum tuberosum]	EST492465 cSTS Solanum tuberosum cDNA clo
459	G2346	BG442540	1.00E-23	[Gossypium arboreum]	GA_Ea0017G06f Gossypium arboreum 7-10 d
459	G2346	AZ919034	1.90E-23	[Zea mays]	1006013G02.x3 1006 - RescueMu Grid G Zea mays geno
459	G2346	BE596165	2.70E-23	[Sorghum bicolor]	PI1_50_D04.b1_A002 Pathogen induced 1 (PI1)
459	G2346	AI443033	2.30E-22	[Glycine max]	sa31a08.y1 Gm-c1004 Glycine max cDNA clone GENO
459	G2346	BF482644	4.30E-22	[Triticum aestivum]	WHE2301-2304_A21_A21ZS Wheat pre-anthesis
459	G2346	gi5931643	6.20E-45	[Antirrhinum majus]	squamosa promoter binding protein-homol
459	G2346	gi5931786	4.20E-26	[Zea mays]	SBP-domain protein 5.
459	G2346	gi8468036	3.30E-14	[Oryza sativa]	Similar to Arabidopsis thaliana chromosome 2
459	G2346	gi9087308	8.30E-08	[Mitochondrion Beta vulgaris var. altissima]	orf102a.
285	G1354	BG128374	2.90E-58	[Lycopersicon esculentum]	EST474020 tomato shoot/meristem Lyc
285	G1354	BE202831	1.90E-56	[Medicago truncatula]	EST402853 KV1 Medicago truncatula cDNA
285	G1354	AI161918	6.60E-55	[Populus tremula x Populus tremuloides]	A009P50U Hybrid aspen
285	G1354	AB028186	1.20E-53	[Oryza sativa]	mRNA for OsNAC7 protein, complete cds.
285	G1354	BE060921	8.00E-50	[Hordeum vulgare]	HVSMEg0013N15f Hordeum vulgare pre-anthesis
285	G1354	AF402603	1.50E-42	[Phaseolus vulgaris]	NAC domain protein NAC2 mRNA, complete c
285	G1354	BE357920	1.60E-42	[Sorghum bicolor]	DG1_23_F03.b1_A002 Dark Grown 1 (DG1) Sorgh
285	G1354	PHRNANAM	3.60E-42	[Petunia x hybrida]	P.hybrida mRNA encoding NAM protein.
285	G1354	AW185617	5.30E-40	[Glycine max]	se80b05.y1 Gm-c1023 Glycine max cDNA clone GENO
285	G1354	gi6006373	4.50E-63	[Oryza sativa]	Similar to NAM like protein (AC005310).
285	G1354	gi15148914	2.30E-44	[Phaseolus vulgaris]	NAC domain protein NAC2.
285	G1354	gi14485513	3.50E-44	[Solanum tuberosum]	putative NAC domain protein.
285	G1354	gi1279640	5.90E-44	[Petunia x hybrida]	NAM.
285	G1354	gi6175246	5.20E-41	[Lycopersicon esculentum]	jasmonic acid 2.

Table 5

285	G1354	gi4218535	5.10E-39	[Triticum sp.]	GRAB1 protein.
285	G1354	gi6732158	5.10E-39	[Triticum monococcum]	unnamed protein product.
285	G1354	gi7716952	3.30E-35	[Medicago truncatula]	NAC1.
285	G1354	gi4996349	2.50E-26	[Nicotiana tabacum]	NAC-domain protein.
285	G1354	gi2982275	3.10E-14	[Picea mariana]	ATAF1-like protein.
119	G1063	BH700922	4.50E-90	[Brassica oleracea]	BOMMZ07TR BO 2 3 KB Brassica oleracea gen
119	G1063	BE451174	2.40E-41	[Lycopersicon esculentum]	EST402062 tomato root, plants pre-a
119	G1063	AW832545	2.00E-40	[Glycine max]	sm12e10.y1 Gm-c1027 Glycine max cDNA clone GENO
119	G1063	AP004693	5.90E-37	[Oryza sativa]	chromosome 8 clone P0461F06, *** SEQUENCING IN
119	G1063	AP004462	4.40E-32	[Oryza sativa (japonica cultivar-group)]	() chromosome 8 clo
119	G1063	AT002234	8.90E-32	[Brassica rapa subsp. pekinensis]	AT002234 Flower bud cDNA Br
119	G1063	BF263465	5.40E-25	[Hordeum vulgare]	HV_CEa0006N02f Hordeum vulgare seedling gre
119	G1063	BG557011	4.20E-22	[Sorghum bicolor]	EM1_41_E02.g1_A002 Embryo 1 (EM1) Sorghum b
119	G1063	BG842856	3.10E-21	[Zea mays]	MEST40-H05.T3 ISUM4-TN Zea mays cDNA clone MEST40-
119	G1063	BG559930	1.40E-18	[Sorghum propinquum]	RHIZ2_75_D09.g1_A003 Rhizome2 (RHIZ2) So
119	G1063	gi15528743	4.20E-26	[Oryza sativa]	contains EST C74560(E31855)-unknown protein.
119	G1063	gi6166283	8.10E-10	[Pinus taeda]	helix-loop-helix protein 1A.
119	G1063	gi11045087	8.80E-09	[Brassica napus]	putative protein.
119	G1063	gi10998404	7.10E-08	[Petunia x hybrida]	anthocyanin 1.
119	G1063	gi99441	2.60E-07	[Volvox carter]	sulfated surface glycoprotein 185 - Volvox
119	G1063	gi1142621	5.00E-07	[Phaseolus vulgaris]	phaseolin G-box binding protein PG2.
119	G1063	gi166428	8.10E-07	[Antirrhinum majus]	DEL.
119	G1063	gi1247386	9.50E-07	[Nicotiana glauca]	PRP2.
119	G1063	gi82091	1.00E-06	[Lycopersicon esculentum]	hydroxyproline-rich glycoprotein
119	G1063	gi1486263	1.40E-06	[Catharanthus roseus]	extensin.
129	G2143	BH650724	3.00E-88	[Brassica oleracea]	BOMIW43TR BO 2 3 KB Brassica oleracea gen
129	G2143	AW832545	1.50E-40	[Glycine max]	sm12e10.y1 Gm-c1027 Glycine max cDNA clone GENO
129	G2143	BE451174	3.50E-40	[Lycopersicon esculentum]	EST402062 tomato root, plants pre-a
129	G2143	AP004693	4.00E-38	[Oryza sativa]	chromosome 8 clone P0461F06, *** SEQUENCING IN
129	G2143	AP004584	6.30E-33	[Oryza sativa (japonica cultivar-group)]	() chromosome 8 clo
129	G2143	AT002234	3.00E-31	[Brassica rapa subsp. pekinensis]	AT002234 Flower bud cDNA Br
129	G2143	BF263465	2.90E-26	[Hordeum vulgare]	HV_CEa0006N02f Hordeum vulgare seedling gre
129	G2143	BG557011	2.60E-22	[Sorghum bicolor]	EM1_41_E02.g1_A002 Embryo 1 (EM1) Sorghum b
129	G2143	BG842856	3.50E-20	[Zea mays]	MEST40-H05.T3 ISUM4-TN Zea mays cDNA clone MEST40-
129	G2143	BG559930	6.10E-18	[Sorghum propinquum]	RHIZ2_75_D09.g1_A003 Rhizome2 (RHIZ2) So
129	G2143	gi15528743	5.50E-26	[Oryza sativa]	contains EST C74560(E31855)-unknown protein.

Table 5

129	G2143	gi1086538	7.60E-09	[Oryza rufipogon]	transcriptional activator Rb homolog.
129	G2143	gi6166283	1.10E-08	[Pinus taeda]	helix-loop-helix protein 1A.
129	G2143	gi1142621	4.60E-07	[Phaseolus vulgaris]	phaseolin G-box binding protein PG2.
129	G2143	gi3339977	5.20E-07	[Glycine max]	symbiotic ammonium transporter; nodulin.
129	G2143	gi5923912	6.10E-07	[Tulipa gesneriana]	bHLH transcription factor GBOF-1.
129	G2143	gi1098404	9.20E-07	[Petunia x hybrida]	anthocyanin 1.
129	G2143	gi4321762	5.20E-06	[Zea mays]	transcription factor MYC7E.
129	G2143	gi166428	6.00E-06	[Antirrhinum majus]	DEL.
129	G2143	gi527665	7.40E-06	[Sorghum bicolor]	myc-like regulatory R gene product.
133	G2557	BH511840	6.70E-62	[Brassica oleracea]	BOGRJ19TR BOGR Brassica oleracea genomic
133	G2557	BE347811	3.70E-46	[Glycine max]	sp05h10.y1 Gm-c1041 Glycine max cDNA clone GENO
133	G2557	AP003141	2.40E-33	[Oryza sativa]	genomic DNA, chromosome 1, PAC clone:P0002B05.
133	G2557	BF263465	3.00E-31	[Hordeum vulgare]	HV_CEa0006N02f Hordeum vulgare seedling gre
133	G2557	AT002234	6.60E-27	[Brassica rapa subsp. pekinensis]	AT002234 Flower bud cDNA Br
133	G2557	BG557011	6.40E-26	[Sorghum bicolor]	EM1_41_E02.g1_A002 Embryo 1 (EM1) Sorghum b
133	G2557	AP004462	7.90E-26	[Oryza sativa (japonica cultivar-group)]	() chromosome 8 clo
133	G2557	BE451174	3.90E-25	[Lycopersicon esculentum]	EST402062 tomato root, plants pre-a
133	G2557	BG842856	5.60E-22	[Zea mays]	MEST40-H05.T3 SUM4-TN Zea mays cDNA clone MEST40-
133	G2557	BG559930	7.00E-14	[Sorghum propinquum]	RHIZ2_75_D09.g1_A003 Rhizome2 (RHIZ2) So
133	G2557	gi15289790	2.40E-36	[Oryza sativa]	contains EST C74560(E31855)-unknown protein.
133	G2557	gi3339977	2.60E-06	[Glycine max]	symbiotic ammonium transporter; nodulin.
133	G2557	gi4206118	1.10E-05	[Mesembryanthemum crystallinum]	transporter homolog.
133	G2557	gi6166283	1.30E-05	[Pinus taeda]	helix-loop-helix protein 1A.
133	G2557	gi527655	3.70E-05	[Pennisetum glaucum]	myc-like regulatory R gene product.
133	G2557	gi5923912	3.70E-05	[Tulipa gesneriana]	bHLH transcription factor GBOF-1.
133	G2557	gi527661	7.80E-05	[Phyllostachys acuta]	myc-like regulatory R gene product.
133	G2557	gi527665	9.50E-05	[Sorghum bicolor]	myc-like regulatory R gene product.
133	G2557	gi1086538	0.0001	[Oryza rufipogon]	transcriptional activator Rb homolog.
133	G2557	gi5669656	0.00013	[Lycopersicon esculentum]	ER33 protein.
697	G2430	BF632520	1.90E-14	[Medicago truncatula]	NF039A08DT1F1054 Drought Medicago trunc
697	G2430	AW396912	1.20E-13	[Glycine max]	sg64g09.y1 Gm-c1007 Glycine max cDNA clone GENO
697	G2430	D41804	4.50E-13	[Oryza sativa]	RICS4626A Rice shoot Oryza sativa cDNA, mRNA s
697	G2430	BE214029	2.60E-10	[Hordeum vulgare]	HV_CE00001P06f Hordeum vulgare seedling gre
697	G2430	AW564570	2.70E-10	[Sorghum bicolor]	LG1_296_E01.b1_A002 Light Grown 1 (LG1) Sor
697	G2430	BG129795	5.40E-10	[Lycopersicon esculentum]	EST475441 tomato shoot/meristem Lyc
697	G2430	AB060130	5.40E-09	[Zea mays]	ZmRRR8 mRNA for response regulator 8, complete cds.

Table 5

697	G2430	BF587105	2.50E-05	[Sorghum propinquum]	FM1_32_C05.b1_A003 Floral-Induced Merist
697	G2430	AI163121	0.3	[Populus tremula x Populus tremuloides]	A033P70U Hybrid aspen
697	G2430	BG595628	0.46	[Solanum tuberosum]	EST494306 cSTS Solanum tuberosum cDNA clo
697	G2430	gi13661174	5.40E-18	[Zea mays]	response regulator 8.
697	G2430	gi15289981	0.028	[Oryza sativa]	hypothetical protein.
697	G2430	gi6942190	0.12	[Mesembryanthemum crystallinum]	CDPK substrate protein 1; C
697	G2430	gi4519671	0.2	[Nicotiana tabacum]	transfactor:
831	G1478	BF275913	1.50E-20	[Gossypium arboreum]	GA_Eb0025C07f Gossypium arboreum 7-10 d
831	G1478	BG157399	6.50E-19	[Glycine max]	sab36g12.y1 Gm-c1026 Glycine max cDNA clone GEN
831	G1478	C95300	2.20E-10	[Citrus unshiu]	C95300 Citrus unshiu Miyagawa-wase maturation
831	G1478	AW034552	2.70E-10	[Lycopersicon esculentum]	EST278168 tomato callus, TAMU Lycop
831	G1478	BI070429	3.40E-10	[Populus tremula x Populus tremuloides]	C037P68U Populus tra
831	G1478	AF016011	5.10E-09	[Brassica napus]	CONSTANS homolog (Bn9CON10) gene, complete c
831	G1478	BE598912	6.20E-09	[Sorghum bicolor]	PI1_84_H11.b1_A002 Pathogen induced 1 (P11)
831	G1478	BG605313	6.80E-09	[Triticum aestivum]	WHE2331_C04_F07ZS Wheat pre-anthesis spik
831	G1478	BE558327	8.90E-09	[Hordeum vulgare]	HV_CE0017D19f Hordeum vulgare seedling gre
831	G1478	BG647091	1.20E-08	[Medicago truncatula]	EST508710 HOGA Medicago truncatula cDNA
831	G1478	gi2895188	4.70E-11	[Brassica napus]	CONSTANS homolog.
831	G1478	gi3618308	1.50E-09	[Oryza sativa]	zinc finger protein.
831	G1478	gi11037308	4.70E-09	[Brassica nigra]	constans-like protein.
831	G1478	gi3341723	1.30E-08	[Raphanus sativus]	CONSTANS-like 1 protein.
831	G1478	gi4091806	1.50E-07	[Malus x domestica]	CONSTANS-like protein 2.
831	G1478	gi10946337	3.10E-07	[Ipomoea nil]	CONSTANS-like protein.
831	G1478	gi4557093	1.40E-05	[Pinus radiata]	zinc finger protein.
831	G1478	gi619312	0.9	[Capparis masakal]	mabinlin III B-chain=sweet protein mabi
831	G1478	gi4732091	1	[Zea mays]	bundle sheath defective protein 2.
831	G1478	gi4699629	1	[Nicotiana glauca]	Chain A, Putative Ancestral Protein Encod
579	G681	BG128147	6.80E-41	[Lycopersicon esculentum]	EST473793 tomato shoot/meristem Lyc
579	G681	BF054497	1.50E-39	[Solanum tuberosum]	EST439727 potato leaves and petioles Sola
579	G681	BE054276	8.40E-39	[Gossypium arboreum]	GA_Ea0002O18f Gossypium arboreum 7-10 d
579	G681	BG269414	4.00E-38	[Mesembryanthemum crystallinum]	LO-3478T3 Ice plant Lambda Un
579	G681	BF620286	7.40E-38	[Hordeum vulgare]	HVSMEE0019F08f Hordeum vulgare seedling sho
579	G681	BE490032	1.00E-37	[Triticum aestivum]	WHE0364_C04_E08ZS Wheat cold-stressed see
579	G681	BI542536	1.40E-36	[Zea mays]	949021A03.y1 949 - Juvenile leaf and shoot cDNA fr
579	G681	BF425254	7.20E-36	[Glycine max]	su42c10.y1 Gm-c1068 Glycine max cDNA clone GENO
579	G681	AW672062	3.20E-34	[Sorghum bicolor]	LG1_354_G05.b1_A002 Light Grown 1 (LG1) Sor

Table 5

579	G681	BG448527	1.00E-33	[Medicago truncatula]	NF036F04RT1F1032 Developing root Medica
579	G681	gi13346188	9.10E-37	[Gossypium hirsutum]	GhMYB25.
579	G681	gi20563	6.30E-36	[Petunia x hybrida]	protein 1.
579	G681	gi485867	1.20E-34	[Antirrhinum majus]	mixta.
579	G681	gi2605617	1.70E-32	[Oryza sativa]	OSMYB1.
579	G681	gi1430846	2.00E-31	[Lycopersicon esculentum]	myb-related transcription factor.
579	G681	gi6651292	2.20E-30	[Pimpinella brachycarpa]	myb-related transcription factor.
579	G681	gi15042116	4.90E-30	[Zea mays subsp. parviglumis]	Cl protein.
579	G681	gi82730	6.10E-30	[Zea mays]	transforming protein (myb) homolog (clone Zm38)
579	G681	gi5139806	8.30E-30	[Glycine max]	GmMYB29A2.
579	G681	gi19055	1.10E-29	[Hordeum vulgare]	MybHV5.
611	G878	AF096299	6.20E-90	[Nicotiana tabacum]	DNA-binding protein 2 (WRKY2) mRNA, compl
611	G878	CUSSLDB	1.80E-83	[Cucumis sativus]	SPF1-like DNA-binding protein mRNA, complet
611	G878	AF193802	3.50E-63	[Oryza sativa]	zinc finger transcription factor WRKY1 mRNA, c
611	G878	AX192162	2.20E-62	[Glycine max]	Sequence 9 from Patent WO0149840.
611	G878	IPBSPF1P	3.80E-58	[Ipomoea batatas]	Sweet potato mRNA for SPF1 protein, complet
611	G878	AFABF1	2.00E-56	[Avena fatua]	A. fatua mRNA for DNA-binding protein (clone ABF
611	G878	LES303343	7.20E-55	[Lycopersicon esculentum]	mRNA for hypothetical protein (ORF
611	G878	AX192164	4.00E-54	[Triticum aestivum]	Sequence 11 from Patent WO0149840.
611	G878	AF080595	2.10E-53	[Pimpinella brachycarpa]	zinc finger protein (ZFP1) mRNA, com
611	G878	PCU48831	2.30E-53	[Petroselinum crispum]	DNA-binding protein WRKY1 mRNA, comple
611	G878	gi4322940	3.30E-128	[Nicotiana tabacum]	DNA-binding protein 2.
611	G878	gi927025	1.10E-109	[Cucumis sativus]	SPF1-like DNA-binding protein.
611	G878	gi6689916	1.50E-74	[Oryza sativa]	zinc finger transcription factor WRKY1.
611	G878	gi484261	1.10E-66	[Ipomoea batatas]	SPF1 protein.
611	G878	gi1159877	2.30E-63	[Avena fatua]	DNA-binding protein.
611	G878	gi13620227	4.60E-63	[Lycopersicon esculentum]	hypothetical protein.
611	G878	gi5917653	1.70E-56	[Petroselinum crispum]	zinc-finger type transcription facto
611	G878	gi4894965	5.00E-56	[Avena sativa]	DNA-binding protein WRKY1.
611	G878	gi3420906	8.70E-56	[Pimpinella brachycarpa]	zinc finger protein; WRKY1.
611	G878	gi13620168	4.20E-22	[Capsella rubella]	hypothetical protein.
47	G374	AP004457	1.20E-73	[Oryza sativa (japonica cultivar-group)]	() chromosome 8 clo
47	G374	AP004693	1.90E-73	[Oryza sativa]	chromosome 8 clone P0461F06, *** SEQUENCING IN
47	G374	BH552835	1.30E-62	[Brassica oleracea]	BOHG756TR BOHG Brassica oleracea genomic
47	G374	BG128229	6.50E-55	[Lycopersicon esculentum]	EST473875 tomato shoot/meristem Lyc
47	G374	BG646959	3.20E-46	[Medicago truncatula]	EST508578 HOGA Medicago truncatula cDNA

Table 5

47	G374	BG890162	8.70E-41	[Solanum tuberosum]	EST516013 cSTD Solanum tuberosum cDNA clo
47	G374	AW179366	6.00E-38	[Zea mays]	618046G06.y1 618 - Inbred Tassel cDNA Library Zea
47	G374	BF473206	1.50E-32	[Triticum aestivum]	WHE0922_G12_M24ZS Wheat 5-15 DAP spike cD
47	G374	AW761011	2.90E-29	[Glycine max]	sl61g11.y1 Gm-c1027 Glycine max cDNA clone GENO
47	G374	AJ436050	1.50E-27	[Hordeum vulgare]	AJ436050 S00007 Hordeum vulgare cDNA clone
47	G374	gi422012	0.8	[Sorghum bicolor]	lipid transfer protein - sorghum (fragmen
47	G374	gi1827893	1	[Zea mays]	Maize Nonspecific Lipid Transfer Protein Complex

Traits of interest

Examples of some of the traits that may be desirable in plants, and that may be provided by transforming the plants with the presently disclosed sequences, are listed in Table 6.

Table 6. Genes, traits and utilities that affect plant characteristics

<u>Trait Category</u>	<u>Traits</u>	Transcription factor genes that <u>impact traits</u>	<u>Utility</u> <u>Gene effect on:</u>
Resistance and tolerance	Salt stress resistance	G22; G196; G226; G303; G312; G325; G353; G482; G545; G801; G867; G884; G922; G926; G1452; G1794; G1820; G1836; G1843; G1863; G2053; G2110; G2140; G2153; G2379; G2701; G2713; G2719; G2789	Germination rate, survivability, yield; extended growth range
	Osmotic stress resistance	G47; G175; G188; G303; G325; G353; G489; G502; G526; G921; G922; G926; G1069; G1089; G1452; G1794; G1930; G2140; G2153; G2379; G2701; G2719; G2789;	Germination rate, survivability, yield
	Cold stress resistance; cold germination	G256; G394; G664; G864; G1322; G2130	Germination, growth, earlier planting
	Tolerance to freezing	G303; G325; G353; G720; G912; G913; G1794; G2053; G2140; G2153; G2379; G2701; G2719; G2789	Survivability, yield, appearance, extended range
	Heat stress resistance	G3; G464; G682; G864; G964;	Germination,

		G1305; G1645; G2130 G2430	growth, later planting
	Drought, low humidity resistance	G303; G325; G353; G720; G912; G926; G1452; G1794; G1820; G1843; G2053; G2140; G2153; G2379; G2583; G2701; G2719; G2789	Survivability, yield, extended range
	Radiation resistance	G1052	Survivability, vigor, appearance
	Decreased herbicide sensitivity	G343; G2133; G2517	Resistant to increased herbicide use
	Increased herbicide sensitivity	G374; G877; G1519	Use as a herbicide target
	Oxidative stress	G477; G789; G1807; G2133; G2517	Improved yield, appearance, reduced senescence
	Light response	G183; G354; G375; G1062; G1322; G1331; G1488; G1494; G1521; G1786; G1794; G2144; G2555;	Germination, growth, development, flowering time
Development, morphology	Overall plant architecture	G24; G27; G31; G33; G47; G147; G156; G160; G182; G187; G195; G196; G211; G221; G237; G280; G342; G352; G357; G358; G360; G362; G364; G365; G367; G373; G377; G396; G431; G447; G479; G546; G546; G551; G578; G580; G596; G615; G617; G620; G625;	Vascular tissues, lignin content; cell wall content; appearance

		G638; G658; G716; G725; G727; G730; G740; G770; G858; G865; G869; G872; G904; G910; G912; G920; G939; G963; G977; G979; G987; G988; G993; G1007; G1010; G1014; G1035; G1046; G1049; G1062; G1069; G1070; G1076; G1089; G1093; G1127; G1131; G1145; G1229; G1246; G1304; G1318; G1320; G1330; G1331; G1352; G1354; G1360; G1364; G1379; G1384; G1399; G1415; G1417; G1442; G1453; G1454; G1459; G1460; G1471; G1475; G1477; G1487; G1487; G1492; G1499; G1499; G1531; G1540; G1543; G1543; G1544; G1548; G1584; G1587; G1588; G1589; G1636; G1642; G1747; G1749; G1749; G1751; G1752; G1763; G1766; G1767; G1778; G1789; G1790; G1791; G1793; G1794; G1795; G1800; G1806; G1811; G1835; G1836; G1838; G1839; G1843; G1853; G1855; G1865; G1881; G1882; G1883; G1884; G1891; G1896; G1898; G1902; G1904; G1906; G1913; G1914; G1925; G1929; G1930; G1954; G1958; G1965; G1976; G2057; G2107; G2133; G2134; G2151; G2154; G2157; G2181;	
--	--	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

		G2290; G2299; G2340; G2340; G2346; G2373; G2376; G2424; G2465; G2505; G2509; G2512; G2513; G2519; G2520; G2533; G2534; G2573; G2589; G2687; G2720; G2787; G2789; G2893	
	Size: increased stature	G189; G1073; G1435; G2430	
	Size: reduced stature or dwarfism	G3; G5; G21; G23; G39; G165; G184; G194; G258; G280; G340; G343; G353; G354; G362; G363; G370; G385; G396; G439; G440; G447; G450; G550; G557; G599; G636; G652; G670; G671; G674; G729; G760; G804; G831; G864; G884; G898; G900; G912; G913; G922; G932; G937; G939; G960; G962; G977; G991; G1000; G1008; G1020; G1023; G1053; G1067; G1075; G1137; G1181; G1198; G1228; G1266; G1267; G1275; G1277; G1309; G1311; G1314; G1317; G1322; G1323; G1326; G1332; G1334; G1367; G1381; G1382; G1386; G1421; G1488; G1494; G1537; G1545; G1560; G1586; G1641; G1652; G1655; G1671; G1750; G1756; G1757; G1782; G1786; G1794; G1839; G1845; G1879; G1886; G1888; G1933; G1939; G1943; G1944; G2011; G2094; G2115;	Ornamental; small stature provides wind resistance; creation of dwarf varieties

		G2130; G2132; G2144; G2145; G2147; G2156; G2294; G2313; G2344; G2431; G2510; G2517; G2521; G2893; G2893	
	Fruit size and number	G362	Biomass, yield, cotton boll fiber density
	Flower structure, inflorescence	G47; G259; G353; G354; G671; G732; G988; G1000; G1063; G1140; G1326; G1449; G1543; G1560; G1587; G1645; G1947; G2108; G2143; G2893	Ornamental horticulture; production of saffron or other edible flowers
	Number and development of trichomes	G225; G226; G247; G362; G585; G634; G676; G682; G1014; G1332; G1452; G1795; G2105	Resistance to pests and desiccation; essential oil production
	Seed size, color, and number	G156; G450; G584; G652; G668; G858; G979; G1040; G1062; G1145; G1255; G1494; G1531; G1534; G1594; G2105; G2114;	Yield
	Root development, modifications	G9; G1482; G1534; G1794; G1852; G2053; G2136; G2140	
	Modifications to root hairs	G225; G226	Nutrient, water uptake, pathogen resistance
	Apical dominance	G559; G732; G1255; G1275; G1411; G1488; G1635; G2452; G2509	Ornamental horticulture
	Branching patterns	G568; G988; G1548	Ornamental horticulture, knot reduction, improved

			windscreen
	Leaf shape, color, modifications	G375; G377; G428; G438; G447; G464; G557; G577; G599; G635; G671; G674; G736; G804; G903; G977; G921; G922; G1038; G1063; G1067; G1073; G1075; G1146; G1152; G1198; G1267; G1269; G1452; G1484; G1586; G1594; G1767; G1786; G1792; G1886; G2059; G2094; G2105; G2113; G2117; G2143; G2144; G2431; G2452; G2465; G2587; G2583; G2724;	Appealing shape or shiny leaves for ornamental agriculture, increased biomass or photosynthesis
	Silique	G1134	Ornamental
	Stem morphology	G47; G438; G671; G748; G988; G1000	Ornamental; digestibility
	Shoot modifications	G390; G391	Ornamental stem bifurcations
Disease, Pathogen Resistance	Bacterial	G211; G347; G367; G418; G525; G545; G578; G1049	Yield, appearance, survivability, extended range
	Fungal	G19; G28; G28; G28; G147; G188; G207; G211; G237; G248; G278; G347; G367; G371; G378; G409; G477; G545; G545; G558; G569; G578; G591; G594; G616; G789; G805; G812; G865; G869; G872; G881; G896; G940; G1047; G1049; G1064; G1084; G1196; G1255; G1266;	Yield, appearance, survivability, extended range

		G1363; G1514; G1756; G1792; G1792; G1792; G1792; G1880; G1919; G1919; G1927; G1927; G1936; G1936; G1950; G2069; G2130; G2380; G2380; G2555	
Nutrients	Increased tolerance to nitrogen-limited soils	G225; G226; G1792	
	Increased tolerance to phosphate-limited soils	G419; G545; G561; G1946	
	Increased tolerance to potassium-limited soils	G561; G911	
Hormonal	Hormone sensitivity	G12; G546; G926; G760; G913; G926; G1062; G1069; G1095; G1134; G1330; G1452; G1666; G1820; G2140; G2789	Seed dormancy, drought tolerance; plant form, fruit ripening
Seed biochemistry	Production of seed prenyl lipids, including tocopherol	G214; G259; G490; G652; G748; G883; G1052; G1328; G1930; G2509; G2520	Antioxidant activity, vitamin E
	Production of seed sterols	G20	Precursors for human steroid hormones; cholesterol modulators
	Production of seed glucosinolates	G353; G484; G674; G1272; G1506; G1897; G1946; G2113; G2117; G2155; G2290; G2340	Defense against insects; putative anticancer activity; undesirable in

			animal feeds
	Modified seed oil content	G162; G162; G180; G192; G241; G265; G286; G291; G427; G509; G519; G561; G567; G590; G818; G849; G892; G961; G974; G1063; G1143; G1190; G1198; G1226; G1229; G1323; G1451; G1471; G1478; G1496; G1526; G1543; G1640; G1644; G1646; G1672; G1677; G1750; G1765; G1777; G1793; G1838; G1902; G1946; G1948; G2059; G2123; G2138; G2139; G2343; G2792; G2830	Vegetable oil production; increased caloric value for animal feeds; lutein content
	Modified seed oil composition	G217; G504; G622; G778; G791; G861; G869; G938; G965; G1417; G2192	Heat stability, digestibility of seed oils
	Modified seed protein content	G162; G226; G241; G371; G427; G509; G567; G597; G732; G849; G865; G892; G963; G988; G1323; G1323; G1419; G1478; G1488; G1634; G1637; G1641; G1644; G1652; G1677; G1777; G1777; G1818; G1820; G1903; G1909; G1946; G1946; G1958; G2059; G2117; G2417; G2509	Reduced caloric value for humans
Leaf biochemistry	Production of flavonoids	G1666*	Ornamental pigment production; pathogen resistance; health

			benefits
	Production of leaf glucosinolates	G264; G353; G484; G652; G674; G681; G1069; G1198; G1322; G1421; G1657; G1794; G1897; G1946; G2115; G2117; G2144; G2155; G2155; G2340; G2512; G2520; G2552	Defense against insects; putative anticancer activity; undesirable in animal feeds
	Production of diterpenes	G229	Induction of enzymes involved in alkaloid biosynthesis
	Production of anthocyanin	G546	Ornamental pigment
	Production of leaf phytosterols, inc. stigmastanol, campesterol	G561; G2131; G2424	Precursors for human steroid hormones; cholesterol modulators
	Leaf fatty acid composition	G214; G377; G861; G962; G975; G987; G1266; G1337; G1399; G1465; G1512; G2136; G2147; G2192	Nutritional value; increase in waxes for disease resistance
	Production of leaf prenyl lipids, including tocopherol	G214; G259; G280; G652; G987; G1543; G2509; G2520	Antioxidant activity, vitamin E
Biochemistry, general	Production of miscellaneous secondary metabolites	G229; G663	
	Sugar, starch, hemicellulose composition,	G158; G211; G211; G237; G242; G274; G598; G1012; G1266; G1309; G1309; G1641; G1765; G1865; G2094; G2094;	Food digestibility, hemicellulose & pectin content; fiber content; plant

		G2589; G2589	tensile strength, wood quality, pathogen resistance, pulp production; tuber starch content
Sugar sensing	Plant response to sugars	G26; G38; G43; G207; G218; G241; G254; G263; G308; G536; G567; G567; G680; G867; G912; G956; G996; G1068; G1225; G1314; G1314; G1337; G1759; G1804; G2153; G2379	Photosynthetic rate, carbohydrate accumulation, biomass production, source-sink relationships, senescence
Growth, Reproduction	Plant growth rate and development	G447; G617; G674; G730; G917; G937; G1035; G1046; G1131; G1425; G1452; G1459; G1492; G1589; G1652; G1879; G1943; G2430; G2431; G2465; G2521	Faster growth, increased biomass or yield, improved appearance; delay in bolting
	Embryo development	G167	
	Seed germination rate	G979; G1792; G2130	Yield
	Plant, seedling vigor	G561; G2346	Survivability, yield
	Senescence; cell death	G571; G636; G878; G1050; G1463; G1749; G1944; G2130; G2155; G2340; G2383	Yield, appearance; response to pathogens;
	Modified fertility	G39; G340; G439; G470; G559; G615; G652; G671; G779; G962; G977; G988; G1000; G1063; G1067; G1075;	Prevents or minimizes escape of the pollen of GMOs

		G1266; G1311; G1321; G1326; G1367; G1386; G1421; G1453; G1471; G1453; G1560; G1594; G1635; G1750; G1947; G2011; G2094; G2113; G2115; G2130; G2143; G2147; G2294; G2510; G2893	
	Early flowering	G147; G157; G180; G183; G183; G184; G185; G208; G227; G294; G390; G390; G390; G391; G391; G427; G427; G490; G565; G590; G592; G720; G789; G865; G898; G898; G989; G989; G1037; G1037; G1142; G1225; G1225; G1226; G1242; G1305; G1305; G1380; G1380; G1480; G1480; G1488; G1494; G1545; G1545; G1649; G1706; G1760; G1767; G1767; G1820; G1841; G1841; G1842; G1843; G1843; G1946; G1946; G2010; G2030; G2030; G2144; G2144; G2295; G2295; G2347; G2348; G2348; G2373; G2373; G2509; G2509; G2555; G2555	Faster generation time; synchrony of flowering; potential for introducing new traits to single variety
	Delayed flowering	G8; G47; G192; G214; G234; G361; G362; G562; G568; G571; G591; G680; G736; G748; G859; G878; G910; G912; G913; G971; G994; G1051; G1052; G1073; G1079; G1335; G1435; G1452; G1478;	Delayed time to pollen production of GMO plants; synchrony of flowering; increased yield

		G1789; G1804; G1865; G1865; G1895; G1900; G2007; G2133; G2155; G2291; G2465	
	Extended flowering phase	G1947	
	Flower and leaf development	G259; G353; G377; G580; G638 G652; G858; G869; G917; G922; G932; G1063; G1075; G1140; G1425; G1452; G1499; G1548; G1645; G1865; G1897; G1933; G2094; G2124; G2140; G2143; G2535; G2557	Ornamental applications; decreased fertility
	Flower abscission	G1897	Ornamental: longer retention of flowers

* When co-expressed with G669 and G663

Significance of modified plant traits

Currently, the existence of a series of maturity groups for different latitudes represents a major barrier to the introduction of new valuable traits. Any trait (e.g. disease resistance) has to be bred into each of the different maturity groups separately, a laborious and costly exercise. The availability of single strain, which could be grown at any latitude, would therefore greatly increase the potential for introducing new traits to crop species such as soybean and cotton.

For many of the traits, listed in Table 6 and below, that may be conferred to plants, a single transcription factor gene may be used to increase or decrease, advance or delay, or improve or prove deleterious to a given trait. For example, overexpression of a transcription factor gene that naturally occurs in a plant may cause early flowering relative to non-transformed or wild-type plants. By knocking out the gene, or suppressing the gene (with, for example, antisense suppression) the plant may experience delayed flowering. Similarly, overexpressing or suppressing one or more genes can impart significant differences in production of plant products,

such as different fatty acid ratios. Thus, suppressing a gene that causes a plant to be more sensitive to cold may improve a plant's tolerance of cold.

Salt stress resistance. Soil salinity is one of the more important variables that determines where a plant may thrive. Salinity is especially important for the successful cultivation of crop plants, particular in many parts of the world that have naturally high soil salt concentrations, or where the soil has been over-utilized. Thus, presently disclosed transcription factor genes that provide increased salt tolerance during germination, the seedling stage, and throughout a plant's life cycle would find particular value for imparting survivability and yield in areas where a particular crop would not normally prosper.

Osmotic stress resistance. Presently disclosed transcription factor genes that confer resistance to osmotic stress may increase germination rate under adverse conditions, which could impact survivability and yield of seeds and plants.

Cold stress resistance. The potential utility of presently disclosed transcription factor genes that increase tolerance to cold is to confer better germination and growth in cold conditions. The germination of many crops is very sensitive to cold temperatures. Genes that would allow germination and seedling vigor in the cold would have highly significant utility in allowing seeds to be planted earlier in the season with a high rate of survivability. Transcription factor genes that confer better survivability in cooler climates allow a grower to move up planting time in the spring and extend the growing season further into autumn for higher crop yields.

Tolerance to freezing. The presently disclosed transcription factor genes that impart tolerance to freezing conditions are useful for enhancing the survivability and appearance of plants conditions or conditions that would otherwise cause extensive cellular damage. Thus, germination of seeds and survival may take place at temperatures significantly below that of the mean temperature required for germination of seeds and survival of non-transformed plants. As with salt tolerance, this has the added benefit of increasing the potential range of a crop plant into regions in which it would otherwise succumb. Cold tolerant transformed plants may also be

planted earlier in the spring or later in autumn, with greater success than with non-transformed plants.

Heat stress tolerance. The germination of many crops is also sensitive to high temperatures. Presently disclosed transcription factor genes that provide increased heat tolerance are generally useful in producing plants that germinate and grow in hot conditions, may find particular use for crops that are planted late in the season, or extend the range of a plant by allowing growth in relatively hot climates.

Drought, low humidity tolerance. Strategies that allow plants to survive in low water conditions may include, for example, reduced surface area or surface oil or wax production. A number of presently disclosed transcription factor genes increase a plant's tolerance to low water conditions and provide the benefits of improved survivability, increased yield and an extended geographic and temporal planting range.

Radiation resistance. Presently disclosed transcription factor genes have been shown to increase lutein production. Lutein, like other xanthophylls such as zeaxanthin and violaxanthin, are important in the protection of plants against the damaging effects of excessive light. Lutein contributes, directly or indirectly, to the rapid rise of non-photochemical quenching in plants exposed to high light. Increased tolerance of field plants to visible and ultraviolet light impacts survivability and vigor, particularly for recent transplants. Also affected are the yield and appearance of harvested plants or plant parts. Crop plants engineered with presently disclosed transcription factor genes that cause the plant to produce higher levels of lutein therefore would have improved photoprotection, leading to less oxidative damage and increase vigor, survivability and higher yields under high light and ultraviolet light conditions.

Decreased herbicide sensitivity. Presently disclosed transcription factor genes that confer resistance or tolerance to herbicides (e.g., glyphosate) may find use in providing means to increase herbicide applications without detriment to desirable plants. This would allow for the increased use of a particular herbicide in a local

environment, with the effect of increased detriment to undesirable species and less harm to transgenic, desirable cultivars.

Increased herbicide sensitivity. Knockouts of a number of the presently disclosed transcription factor genes have been shown to be lethal to developing embryos. Thus, these genes are potentially useful as herbicide targets.

Oxidative stress. In plants, as in all living things, abiotic and biotic stresses induce the formation of oxygen radicals, including superoxide and peroxide radicals. This has the effect of accelerating senescence, particularly in leaves, with the resulting loss of yield and adverse effect on appearance. Generally, plants that have the highest level of defense mechanisms, such as, for example, polyunsaturated moieties of membrane lipids, are most likely to thrive under conditions that introduce oxidative stress (e.g., high light, ozone, water deficit, particularly in combination). Introduction of the presently disclosed transcription factor genes that increase the level of oxidative stress defense mechanisms would provide beneficial effects on the yield and appearance of plants. One specific oxidizing agent, ozone, has been shown to cause significant foliar injury, which impacts yield and appearance of crop and ornamental plants. In addition to reduced foliar injury that would be found in ozone resistant plant created by transforming plants with some of the presently disclosed transcription factor genes, the latter have also been shown to have increased chlorophyll fluorescence (Yu-Sen Chang et al. Bot. Bull. Acad. Sin. (2001) 42: 265-272).

Heavy metal tolerance. Heavy metals such as lead, mercury, arsenic, chromium and others may have a significant adverse impact on plant respiration. Plants that have been transformed with presently disclosed transcription factor genes that confer improved resistance to heavy metals, through, for example, sequestering or reduced uptake of the metals will show improved vigor and yield in soils with relatively high concentrations of these elements. Conversely, transgenic transcription factors may also be introduced into plants to confer an increase in heavy metal uptake, which may benefit efforts to clean up contaminated soils.

Light response. Presently disclosed transcription factor genes that modify a plant's response to light may be useful for modifying a plant's growth or

development, for example, photomorphogenesis in poor light, or accelerating flowering time in response to various light intensities, quality or duration to which a non-transformed plant would not similarly respond. Examples of such responses that have been demonstrated include leaf number and arrangement, and early flower bud appearances.

Overall plant architecture. Several presently disclosed transcription factor genes have been introduced into plants to alter numerous aspects of the plant's morphology. For example, it has been demonstrated that a number of transcription factors may be used to manipulate branching, such as the means to modify lateral branching, a possible application in the forestry industry. Transgenic plants have also been produced that have altered cell wall content, lignin production, flower organ number, or overall shape of the plants. Presently disclosed transcription factor genes transformed into plants may be used to affect plant morphology by increasing or decreasing internode distance, both of which may be advantageous under different circumstances. For example, for fast growth of woody plants to provide more biomass, or fewer knots, increased internode distances are generally desirable. For improved wind screening of shrubs or trees, or harvesting characteristics of, for example, members of the Gramineae family, decreased internode distance may be advantageous. These modifications would also prove useful in the ornamental horticulture industry for the creation of unique phenotypic characteristics of ornamental plants.

Increased stature. For some ornamental plants, the ability to provide larger varieties may be highly desirable. For many plants, including fruit-bearing trees or trees and shrubs that serve as view or wind screens, increased stature provides obvious benefits. Crop species may also produce higher yields on larger cultivars.

Reduced stature or dwarfism. Presently disclosed transcription factor genes that decrease plant stature can be used to produce plants that are more resistant to damage by wind and rain, or more resistant to heat or low humidity or water deficit. Dwarf plants are also of significant interest to the ornamental horticulture industry, and particularly for home garden applications for which space availability may be limited.

Fruit size and number. Introduction of presently disclosed transcription factor genes that affect fruit size will have desirable impacts on fruit size and number, which may comprise increases in yield for fruit crops, or reduced fruit yield, such as when vegetative growth is preferred (e.g., with bushy ornamentals, or where fruit is undesirable, as with ornamental olive trees).

Flower structure, inflorescence, and development. Presently disclosed transgenic transcription factors have been used to create plants with larger flowers or arrangements of flowers that are distinct from wild-type or non-transformed cultivars. This would likely have the most value for the ornamental horticulture industry, where larger flowers or interesting presentations generally are preferred and command the highest prices. Flower structure may have advantageous effects on fertility, and could be used, for example, to decrease fertility by the absence, reduction or screening of reproductive components. One interesting application for manipulation of flower structure, for example, by introduced transcription factors could be in the increased production of edible flowers or flower parts, including saffron, which is derived from the stigmas of *Crocus sativus*.

Number and development of trichomes. Several presently disclosed transcription factor genes have been used to modify trichome number and amount of trichome products in plants. Trichome glands on the surface of many higher plants produce and secrete exudates that give protection from the elements and pests such as insects, microbes and herbivores. These exudates may physically immobilize insects and spores, may be insecticidal or ant-microbial or they may act as allergens or irritants to protect against herbivores. Trichomes have also been suggested to decrease transpiration by decreasing leaf surface air flow, and by exuding chemicals that protect the leaf from the sun.

Seed size, color and number. The introduction of presently disclosed transcription factor genes into plants that alter the size or number of seeds may have a significant impact on yield, both when the product is the seed itself, or when biomass of the vegetative portion of the plant is increased by reducing seed production. In the case of fruit products, it is often advantageous to modify a plant to have reduced size

or number of seeds relative to non-transformed plants to provide seedless or varieties with reduced numbers or smaller seeds. Presently disclosed transcription factor genes have also been shown to affect seed size, including the development of larger seeds. Seed size, in addition to seed coat integrity, thickness and permeability, seed water content and by a number of other components including antioxidants and oligosaccharides, may affect seed longevity in storage. This would be an important utility when the seed of a plant is the harvested crops, as with, for example, peas, beans, nuts, etc. Presently disclosed transcription factor genes have also been used to modify seed color, which could provide added appeal to a seed product.

Root development, modifications. By modifying the structure or development of roots by transforming into a plant one or more of the presently disclosed transcription factor genes, plants may be produced that have the capacity to thrive in otherwise unproductive soils. For example, grape roots that extend further into rocky soils, or that remain viable in waterlogged soils, would increase the effective planting range of the crop. It may be advantageous to manipulate a plant to produce short roots, as when a soil in which the plant will be growing is occasionally flooded, or when pathogenic fungi or disease-causing nematodes are prevalent.

Modifications to root hairs. Presently disclosed transcription factor genes that increase root hair length or number potentially could be used to increase root growth or vigor, which might in turn allow better plant growth under adverse conditions such as limited nutrient or water availability.

Apical dominance. The modified expression of presently disclosed transcription factors that control apical dominance could be used in ornamental horticulture, for example, to modify plant architecture.

Branching patterns. Several presently disclosed transcription factor genes have been used to manipulate branching, which could provide benefits in the forestry industry. For example, reduction in the formation of lateral branches could reduce knot formation. Conversely, increasing the number of lateral branches could provide utility when a plant is used as a windscreen, or may also provide ornamental advantages.

Leaf shape, color and modifications. It has been demonstrated in laboratory experiments that overexpression of some of the presently disclosed transcription factors produced marked effects on leaf development. At early stages of growth, these transgenic seedlings developed narrow, upward pointing leaves with long petioles, possibly indicating a disruption in circadian-clock controlled processes or nyctinastic movements. Other transcription factor genes can be used to increase plant biomass; large size would be useful in crops where the vegetative portion of the plant is the marketable portion.

Siliques. Genes that later silique conformation in brassicates may be used to modify fruit ripening processes in brassicates and other plants, which may positively affect seed or fruit quality.

Stem morphology and shoot modifications. Laboratory studies have demonstrated that introducing several of the presently disclosed transcription factor genes into plants can cause stem bifurcations in shoots, in which the shoot meristems split to form two or three separate shoots. This unique appearance would be desirable in ornamental applications.

Diseases, pathogens and pests. A number of the presently disclosed transcription factor genes have been shown to or are likely to confer resistance to various plant diseases, pathogens and pests. The offending organisms include fungal pathogens *Fusarium oxysporum*, *Botrytis cinerea*, *Sclerotinia sclerotiorum*, and *Erysiphe orontii*. Bacterial pathogens to which resistance may be conferred include *Pseudomonas syringae*. Other problem organisms may potentially include nematodes, mollicutes, parasites, or herbivorous arthropods. In each case, one or more transformed transcription factor genes may provide some benefit to the plant to help prevent or overcome infestation. The mechanisms by which the transcription factors work could include increasing surface waxes or oils, surface thickness, local senescence, or the activation of signal transduction pathways that regulate plant defense in response to attacks by herbivorous pests (including, for example, protease inhibitors).

Increased tolerance of plants to nutrient-limited soils. Presently disclosed transcription factor genes introduced into plants may provide the means to improve uptake of essential nutrients, including nitrogenous compounds, phosphates, potassium, and trace minerals. The effect of these modifications is to increase the seedling germination and range of ornamental and crop plants. The utilities of presently disclosed transcription factor genes conferring tolerance to conditions of low nutrients also include cost savings to the grower by reducing the amounts of fertilizer needed, environmental benefits of reduced fertilizer runoff; and improved yield and stress tolerance. In addition, this gene could be used to alter seed protein amounts and/or composition that could impact yield as well as the nutritional value and production of various food products.

Hormone sensitivity. One or more of the presently disclosed transcription factor genes have been shown to affect plant abscisic acid (ABA) sensitivity. This plant hormone is likely the most important hormone in mediating the adaptation of a plant to stress. For example, ABA mediates conversion of apical meristems into dormant buds. In response to increasingly cold conditions, the newly developing leaves growing above the meristem become converted into stiff bud scales that closely wrap the meristem and protect it from mechanical damage during winter. ABA in the bud also enforces dormancy; during premature warm spells, the buds are inhibited from sprouting. Bud dormancy is eliminated after either a prolonged cold period of cold or a significant number of lengthening days. Thus, by affecting ABA sensitivity, introduced transcription factor genes may affect cold sensitivity and survivability. ABA is also important in protecting plants from drought tolerance.

Several other of the present transcription factor genes have been used to manipulate ethylene signal transduction and response pathways. These genes can thus be used to manipulate the processes influenced by ethylene, such as seed germination or fruit ripening, and to improve seed or fruit quality.

Production of seed and leaf prenyl lipids, including tocopherol. Prenyl lipids play a role in anchoring proteins in membranes or membranous organelles. Thus modifying the prenyl lipid content of seeds and leaves could affect membrane integrity and function. A number of presently disclosed transcription factor genes

have been shown to modify the tocopherol composition of plants. Tocopherols have both anti-oxidant and vitamin E activity.

Production of seed and leaf phytosterols: Presently disclosed transcription factor genes that modify levels of phytosterols in plants may have at least two utilities. First, phytosterols are an important source of precursors for the manufacture of human steroid hormones. Thus, regulation of transcription factor expression or activity could lead to elevated levels of important human steroid precursors for steroid semi-synthesis. For example, transcription factors that cause elevated levels of campesterol in leaves, or sitosterols and stigmasterols in seed crops, would be useful for this purpose. Phytosterols and their hydrogenated derivatives phytostanols also have proven cholesterol-lowering properties, and transcription factor genes that modify the expression of these compounds in plants would thus provide health benefits.

Production of seed and leaf glucosinolates. Some glucosinolates have anti-cancer activity; thus, increasing the levels or composition of these compounds by introducing several of the presently disclosed transcription factors might be of interest from a nutraceutical standpoint. (3) Glucosinolates form part of a plants natural defense against insects. Modification of glucosinolate composition or quantity could therefore afford increased protection from predators. Furthermore, in edible crops, tissue specific promoters might be used to ensure that these compounds accumulate specifically in tissues, such as the epidermis, which are not taken for consumption.

Modified seed oil content. The composition of seeds, particularly with respect to seed oil amounts and/or composition, is very important for the nutritional value and production of various food and feed products. Several of the presently disclosed transcription factor genes in seed lipid saturation that alter seed oil content could be used to improve the heat stability of oils or to improve the nutritional quality of seed oil, by, for example, reducing the number of calories in seed, increasing the number of calories in animal feeds, or altering the ratio of saturated to unsaturated lipids comprising the oils.

Seed and leaf fatty acid composition. A number of the presently disclosed transcription factor genes have been shown to alter the fatty acid composition in plants, and seeds in particular. This modification may find particular value for improving the nutritional value of, for example, seeds or whole plants. Dietary fatty acids ratios have been shown to have an effect on, for example, bone integrity and remodeling (see, for example, Weiler, H.A., *Pediatr Res* (2000) 47:5 692-697). The ratio of dietary fatty acids may alter the precursor pools of long-chain polyunsaturated fatty acids that serve as precursors for prostaglandin synthesis. In mammalian connective tissue, prostaglandins serve as important signals regulating the balance between resorption and formation in bone and cartilage. Thus dietary fatty acid ratios altered in seeds may affect the etiology and outcome of bone loss.

Modified seed protein content. As with seed oils, the composition of seeds, particularly with respect to protein amounts and/or composition, is very important for the nutritional value and production of various food and feed products. A number of the presently disclosed transcription factor genes modify the protein concentrations in seeds would provide nutritional benefits, and may be used to prolong storage, increase seed pest or disease resistance, or modify germination rates.

Production of flavonoids in leaves and other plant parts. Expression of presently disclosed transcription factor genes that increase flavonoid production in plants, including anthocyanins and condensed tannins, may be used to alter in pigment production for horticultural purposes, and possibly increasing stress resistance. Flavonoids have antimicrobial activity and could be used to engineer pathogen resistance. Several flavonoid compounds have health promoting effects such as the inhibition of tumor growth and cancer, prevention of bone loss and the prevention of the oxidation of lipids. Increasing levels of condensed tannins, whose biosynthetic pathway is shared with anthocyanin biosynthesis, in forage legumes is an important agronomic trait because they prevent pasture bloat by collapsing protein foams within the rumen. For a review on the utilities of flavonoids and their derivatives, refer to Dixon et al. (1999) *Trends Plant Sci.* 4:394-400.

Production of diterpenes in leaves and other plant parts. Depending on the plant species, varying amounts of diverse secondary biochemicals (often lipophilic

terpenes) are produced and exuded or volatilized by trichomes. These exotic secondary biochemicals, which are relatively easy to extract because they are on the surface of the leaf, have been widely used in such products as flavors and aromas, drugs, pesticides and cosmetics. Thus, the overexpression of genes that are used to produce diterpenes in plants may be accomplished by introducing transcription factor genes that induce said overexpression. One class of secondary metabolites, the diterpenes, can effect several biological systems such as tumor progression, prostaglandin synthesis and tissue inflammation. In addition, diterpenes can act as insect pheromones, termite allomones, and can exhibit neurotoxic, cytotoxic and antimitotic activities. As a result of this functional diversity, diterpenes have been the target of research several pharmaceutical ventures. In most cases where the metabolic pathways are impossible to engineer, increasing trichome density or size on leaves may be the only way to increase plant productivity.

Production of anthocyanin in leaves and other plant parts. Several presently disclosed transcription factor genes can be used to alter anthocyanin production in numerous plant species. The potential utilities of these genes include alterations in pigment production for horticultural purposes, and possibly increasing stress resistance in combination with another transcription factor.

Production of miscellaneous secondary metabolites. Microarray data suggests that flux through the aromatic amino acid biosynthetic pathways and primary and secondary metabolite biosynthetic pathways are up-regulated. Presently disclosed transcription factors have been shown to be involved in regulating alkaloid biosynthesis, in part by up-regulating the enzymes indole-3-glycerol phosphatase and strictosidine synthase. Phenylalanine ammonia lyase, chalcone synthase and trans-cinnamate mono-oxygenase are also induced, and are involved in phenylpropanoid biosynthesis.

Sugar, starch, hemicellulose composition. Overexpression of the presently disclosed transcription factors that affect sugar content resulted in plants with altered leaf insoluble sugar content. Transcription factors that alter plant cell wall composition have several potential applications including altering food digestibility, plant tensile strength, wood quality, pathogen resistance and in pulp production. The

potential utilities of a gene involved in glucose-specific sugar sensing are to alter energy balance, photosynthetic rate, carbohydrate accumulation, biomass production, source-sink relationships, and senescence.

Hemicellulose is not desirable in paper pulps because of its lack of strength compared with cellulose. Thus modulating the amounts of cellulose vs. hemicellulose in the plant cell wall is desirable for the paper/lumber industry. Increasing the insoluble carbohydrate content in various fruits, vegetables, and other edible consumer products will result in enhanced fiber content. Increased fiber content would not only provide health benefits in food products, but might also increase digestibility of forage crops. In addition, the hemicellulose and pectin content of fruits and berries affects the quality of jam and catsup made from them. Changes in hemicellulose and pectin content could result in a superior consumer product.

Plant response to sugars and sugar composition. In addition to their important role as an energy source and structural component of the plant cell, sugars are central regulatory molecules that control several aspects of plant physiology, metabolism and development. It is thought that this control is achieved by regulating gene expression and, in higher plants, sugars have been shown to repress or activate plant genes involved in many essential processes such as photosynthesis, glyoxylate metabolism, respiration, starch and sucrose synthesis and degradation, pathogen response, wounding response, cell cycle regulation, pigmentation, flowering and senescence. The mechanisms by which sugars control gene expression are not understood.

Because sugars are important signaling molecules, the ability to control either the concentration of a signaling sugar or how the plant perceives or responds to a signaling sugar could be used to control plant development, physiology or metabolism. For example, the flux of sucrose (a disaccharide sugar used for systemically transporting carbon and energy in most plants) has been shown to affect gene expression and alter storage compound accumulation in seeds. Manipulation of the sucrose signaling pathway in seeds may therefore cause seeds to have more protein, oil or carbohydrate, depending on the type of manipulation. Similarly, in tubers, sucrose is converted to starch which is used as an energy store. It is thought that sugar signaling pathways may partially determine the levels of starch synthesized

in the tubers. The manipulation of sugar signaling in tubers could lead to tubers with a higher starch content.

Thus, the presently disclosed transcription factor genes that manipulate the sugar signal transduction pathway may lead to altered gene expression to produce plants with desirable traits. In particular, manipulation of sugar signal transduction pathways could be used to alter source-sink relationships in seeds, tubers, roots and other storage organs leading to increase in yield.

Plant growth rate and development. A number of the presently disclosed transcription factor genes have been shown to have significant effects on plant growth rate and development. These observations have included, for example, more rapid or delayed growth and development of reproductive organs. This would provide utility for regions with short or long growing seasons, respectively. Accelerating plant growth would also improve early yield or increase biomass at an earlier stage, when such is desirable (for example, in producing forestry products).

Embryo development. Presently disclosed transcription factor genes that alter embryo development has been used to alter seed protein and oil amounts and/or composition which is very important for the nutritional value and production of various food products. Seed shape and seed coat may also be altered by these genes, which may provide for improved storage stability.

Seed germination rate. A number of the presently disclosed transcription factor genes have been shown to modify seed germination rate, including when the seeds are in conditions normally unfavorable for germination (e.g., cold, heat or salt stress, or in the presence of ABA), and may thus be used to modify and improve germination rates under adverse conditions.

Plant, seedling vigor. Seedlings transformed with presently disclosed transcription factors have been shown to possess larger cotyledons and appeared somewhat more advanced than control plants. This indicates that the seedlings developed more rapidly than the control plants. Rapid seedling development is likely to reduce loss due to diseases particularly prevalent at the seedling stage (e.g.,

damping off) and is thus important for survivability of plants germinating in the field or in controlled environments.

Senescence, cell death. Presently disclosed transcription factor genes may be used to alter senescence responses in plants. Although leaf senescence is thought to be an evolutionary adaptation to recycle nutrients, the ability to control senescence in an agricultural setting has significant value. For example, a delay in leaf senescence in some maize hybrids is associated with a significant increase in yields and a delay of a few days in the senescence of soybean plants can have a large impact on yield. Delayed flower senescence may also generate plants that retain their blossoms longer and this may be of potential interest to the ornamental horticulture industry.

Modified fertility. Plants that overexpress a number of the presently disclosed transcription factor genes have been shown to possess reduced fertility. This could be a desirable trait, as it could be exploited to prevent or minimize the escape of the pollen of genetically modified organisms (GMOs) into the environment.

Early and delayed flowering. Presently disclosed transcription factor genes that accelerate flowering could have valuable applications in such programs since they allow much faster generation times. In a number of species, for example, broccoli, cauliflower, where the reproductive parts of the plants constitute the crop and the vegetative tissues are discarded, it would be advantageous to accelerate time to flowering. Accelerating flowering could shorten crop and tree breeding programs. Additionally, in some instances, a faster generation time might allow additional harvests of a crop to be made within a given growing season. A number of *Arabidopsis* genes have already been shown to accelerate flowering when constitutively expressed. These include LEAFY, APETALA1 and CONSTANS (Mandel, M. et al., 1995, Nature 377, 522-524; Weigel, D. and Nilsson, O., 1995, Nature 377, 495-500; Simon et al., 1996, Nature 384, 59-62).

By regulating the expression of potential flowering using inducible promoters, flowering could be triggered by application of an inducer chemical. This would allow flowering to be synchronized across a crop and facilitate more efficient harvesting. Such inducible systems could also be used to tune the flowering of crop varieties to

different latitudes. At present, species such as soybean and cotton are available as a series of maturity groups that are suitable for different latitudes on the basis of their flowering time (which is governed by day-length). A system in which flowering could be chemically controlled would allow a single high-yielding northern maturity group to be grown at any latitude. In southern regions such plants could be grown for longer, thereby increasing yields, before flowering was induced. In more northern areas, the induction would be used to ensure that the crop flowers prior to the first winter frosts.

In a sizeable number of species, for example, root crops, where the vegetative parts of the plants constitute the crop and the reproductive tissues are discarded, it would be advantageous to delay or prevent flowering. Extending vegetative development with presently disclosed transcription factor genes could thus bring about large increases in yields.. Prevention of flowering might help maximize vegetative yields and prevent escape of genetically modified organism (GMO) pollen.

Extended flowering phase. Presently disclosed transcription factors that extend flowering time have utility in engineering plants with longer-lasting flowers for the horticulture industry, and for extending the time in which the plant is fertile.

Flower and leaf development. Presently disclosed transcription factor genes have been used to modify the development of flowers and leaves. This could be advantageous in the development of new ornamental cultivars that present unique configurations. In addition, some of these genes have been shown to reduce a plant's fertility, which is also useful for helping to prevent development of pollen of GMOs.

Flower abscission. Presently disclosed transcription factor genes introduced into plants have been used to retain flowers for longer periods. This would provide a significant benefit to the ornamental industry, for both cut flowers and woody plant varieties (of, for example, maize), as well as have the potential to lengthen the fertile period of a plant, which could positively impact yield and breeding programs.

A listing of specific effects and utilities that the presently disclosed transcription factor genes have on plants, as determined by direct observation and assay analysis, is provided in Table 4.

XVI. Antisense and Co-suppression

In addition to expression of the nucleic acids of the invention as gene replacement or plant phenotype modification nucleic acids, the nucleic acids are also useful for sense and anti-sense suppression of expression, e.g., to down-regulate expression of a nucleic acid of the invention, e.g., as a further mechanism for modulating plant phenotype. That is, the nucleic acids of the invention, or subsequences or anti-sense sequences thereof, can be used to block expression of naturally occurring homologous nucleic acids. A variety of sense and anti-sense technologies are known in the art, e.g., as set forth in Lichtenstein and Nellen (1997) Antisense Technology: A Practical Approach IRL Press at Oxford University Press, Oxford, U.K.. In general, sense or anti-sense sequences are introduced into a cell, where they are optionally amplified, e.g., by transcription. Such sequences include both simple oligonucleotide sequences and catalytic sequences such as ribozymes.

For example, a reduction or elimination of expression (i.e., a "knock-out") of a transcription factor or transcription factor homologue polypeptide in a transgenic plant, e.g., to modify a plant trait, can be obtained by introducing an antisense construct corresponding to the polypeptide of interest as a cDNA. For antisense suppression, the transcription factor or homologue cDNA is arranged in reverse orientation (with respect to the coding sequence) relative to the promoter sequence in the expression vector. The introduced sequence need not be the full length cDNA or gene, and need not be identical to the cDNA or gene found in the plant type to be transformed. Typically, the antisense sequence need only be capable of hybridizing to the target gene or RNA of interest. Thus, where the introduced sequence is of shorter length, a higher degree of homology to the endogenous transcription factor sequence will be needed for effective antisense suppression. While antisense sequences of various lengths can be utilized, preferably, the introduced antisense sequence in the vector will be at least 30 nucleotides in length, and improved antisense suppression will typically be observed as the length of the antisense sequence increases. Preferably, the length of the antisense sequence in the vector will be greater than 100 nucleotides.

Transcription of an antisense construct as described results in the production of RNA molecules that are the reverse complement of mRNA molecules transcribed from the endogenous transcription factor gene in the plant cell.

Suppression of endogenous transcription factor gene expression can also be achieved using a ribozyme. Ribozymes are RNA molecules that possess highly specific endoribonuclease activity. The production and use of ribozymes are disclosed in U.S. Patent No. 4,987,071 and U.S. Patent No. 5,543,508. Synthetic ribozyme sequences including antisense RNAs can be used to confer RNA cleaving activity on the antisense RNA, such that endogenous mRNA molecules that hybridize to the antisense RNA are cleaved, which in turn leads to an enhanced antisense inhibition of endogenous gene expression.

Suppression of endogenous transcription factor gene expression can also be achieved using RNA interference, or RNAi. RNAi is a post-transcriptional, targeted gene-silencing technique that uses double-stranded RNA (dsRNA) to incite degradation of messenger RNA (mRNA) containing the same sequence as the dsRNA (Constans, (2002) *The Scientist* 16:36). Small interfering RNAs, or siRNAs are produced in at least two steps: an endogenous ribonuclease cleaves longer dsRNA into shorter, 21-23 nucleotide-long RNAs. The siRNA segments then mediate the degradation of the target mRNA (Zamore, (2001) *Nature Struct. Biol.*, 8:746-50). RNAi has been used for gene function determination in a manner similar to antisense oligonucleotides (Constans, (2002) *The Scientist* 16:36). Expression vectors that continually express siRNAs in transiently and stably transfected have been engineered to express small hairpin RNAs (shRNAs), which get processed in vivo into siRNA-like molecules capable of carrying out gene-specific silencing (Brummelkamp et al., (2002) *Science* 296:550-553, and Paddison, et al. (2002) *Genes & Dev.* 16:948-958). Post-transcriptional gene silencing by double-stranded RNA is discussed in further detail by Hammond et al. (2001) *Nature Rev Gen* 2: 110-119, Fire et al. (1998) *Nature* 391: 806-811 and Timmons and Fire (1998) *Nature* 395: 854.

Vectors in which RNA encoded by a transcription factor or transcription factor homologue cDNA is over-expressed can also be used to obtain co-suppression of a corresponding endogenous gene, e.g., in the manner described in U.S. Patent No.

5,231,020 to Jorgensen. Such co-suppression (also termed sense suppression) does not require that the entire transcription factor cDNA be introduced into the plant cells, nor does it require that the introduced sequence be exactly identical to the endogenous transcription factor gene of interest. However, as with antisense suppression, the suppressive efficiency will be enhanced as specificity of hybridization is increased, e.g., as the introduced sequence is lengthened, and/or as the sequence similarity between the introduced sequence and the endogenous transcription factor gene is increased.

Vectors expressing an untranslatable form of the transcription factor mRNA, e.g., sequences comprising one or more stop codon, or nonsense mutation) can also be used to suppress expression of an endogenous transcription factor, thereby reducing or eliminating its activity and modifying one or more traits. Methods for producing such constructs are described in U.S. Patent No. 5,583,021. Preferably, such constructs are made by introducing a premature stop codon into the transcription factor gene. Alternatively, a plant trait can be modified by gene silencing using double-strand RNA (Sharp (1999) Genes and Development 13: 139-141). Another method for abolishing the expression of a gene is by insertion mutagenesis using the T-DNA of *Agrobacterium tumefaciens*. After generating the insertion mutants, the mutants can be screened to identify those containing the insertion in a transcription factor or transcription factor homologue gene. Plants containing a single transgene insertion event at the desired gene can be crossed to generate homozygous plants for the mutation. Such methods are well known to those of skill in the art. (See for example Koncz et al. (1992) Methods in Arabidopsis Research, World Scientific.)

Alternatively, a plant phenotype can be altered by eliminating an endogenous gene, such as a transcription factor or transcription factor homologue, e.g., by homologous recombination (Kempin et al. (1997) Nature 389:802-803).

A plant trait can also be modified by using the Cre-lox system (for example, as described in US Pat. No. 5,658,772). A plant genome can be modified to include first and second lox sites that are then contacted with a Cre recombinase. If the lox sites are in the same orientation, the intervening DNA sequence between the two sites

is excised. If the lox sites are in the opposite-orientation, the intervening sequence is inverted.

The polynucleotides and polypeptides of this invention can also be expressed in a plant in the absence of an expression cassette by manipulating the activity or expression level of the endogenous gene by other means. For example, by ectopically expressing a gene by T-DNA activation tagging (Ichikawa et al. (1997) Nature 390 698-701; Kakimoto et al. (1996) Science 274: 982-985). This method entails transforming a plant with a gene tag containing multiple transcriptional enhancers and once the tag has inserted into the genome, expression of a flanking gene coding sequence becomes deregulated. In another example, the transcriptional machinery in a plant can be modified so as to increase transcription levels of a polynucleotide of the invention (*See, e.g.,* PCT Publications WO 96/06166 and WO 98/53057 which describe the modification of the DNA-binding specificity of zinc finger proteins by changing particular amino acids in the DNA-binding motif).

The transgenic plant can also include the machinery necessary for expressing or altering the activity of a polypeptide encoded by an endogenous gene, for example by altering the phosphorylation state of the polypeptide to maintain it in an activated state.

Transgenic plants (or plant cells, or plant explants, or plant tissues) incorporating the polynucleotides of the invention and/or expressing the polypeptides of the invention can be produced by a variety of well established techniques as described above. Following construction of a vector, most typically an expression cassette, including a polynucleotide, e.g., encoding a transcription factor or transcription factor homologue, of the invention, standard techniques can be used to introduce the polynucleotide into a plant, a plant cell, a plant explant or a plant tissue of interest. Optionally, the plant cell, explant or tissue can be regenerated to produce a transgenic plant.

The plant can be any higher plant, including gymnosperms, monocotyledonous and dicotyledonous plants. Suitable protocols are available for *Leguminosae* (alfalfa, soybean, clover, etc.), *Umbelliferae* (carrot, celery, parsnip),

Cruciferae (cabbage, radish, rapeseed, broccoli, etc.), *Curcubitaceae* (melons and cucumber), *Gramineae* (wheat, corn, rice, barley, millet, etc.), *Solanaceae* (potato, tomato, tobacco, peppers, etc.), and various other crops. See protocols described in Ammirato et al. (1984) Handbook of Plant Cell Culture –Crop Species, Macmillan Publ. Co. Shimamoto et al. (1989) Nature 338:274-276; Fromm et al. (1990) Bio/Technology 8:833-839; and Vasil et al. (1990) Bio/Technology 8:429-434.

Transformation and regeneration of both monocotyledonous and dicotyledonous plant cells is now routine, and the selection of the most appropriate transformation technique will be determined by the practitioner. The choice of method will vary with the type of plant to be transformed; those skilled in the art will recognize the suitability of particular methods for given plant types. Suitable methods can include, but are not limited to: electroporation of plant protoplasts; liposome-mediated transformation; polyethylene glycol (PEG) mediated transformation; transformation using viruses; micro-injection of plant cells; micro-projectile bombardment of plant cells; vacuum infiltration; and *Agrobacterium tumefaciens* mediated transformation. Transformation means introducing a nucleotide sequence into a plant in a manner to cause stable or transient expression of the sequence.

Successful examples of the modification of plant characteristics by transformation with cloned sequences which serve to illustrate the current knowledge in this field of technology, and which are herein incorporated by reference, include: U.S. Patent Nos. 5,571,706; 5,677,175; 5,510,471; 5,750,386; 5,597,945; 5,589,615; 5,750,871; 5,268,526; 5,780,708; 5,538,880; 5,773,269; 5,736,369 and 5,610,042.

Following transformation, plants are preferably selected using a dominant selectable marker incorporated into the transformation vector. Typically, such a marker will confer antibiotic or herbicide resistance on the transformed plants, and selection of transformants can be accomplished by exposing the plants to appropriate concentrations of the antibiotic or herbicide.

After transformed plants are selected and grown to maturity, those plants showing a modified trait are identified. The modified trait can be any of those traits described above. Additionally, to confirm that the modified trait is due to changes in

expression levels or activity of the polypeptide or polynucleotide of the invention can be determined by analyzing mRNA expression using Northern blots, RT-PCR or microarrays, or protein expression using immunoblots or Western blots or gel shift assays.

XVII. Integrated Systems – Sequence Identity

Additionally, the present invention may be an integrated system, computer or computer readable medium that comprises an instruction set for determining the identity of one or more sequences in a database. In addition, the instruction set can be used to generate or identify sequences that meet any specified criteria. Furthermore, the instruction set may be used to associate or link certain functional benefits, such improved characteristics, with one or more identified sequence.

For example, the instruction set can include, e.g., a sequence comparison or other alignment program, e.g., an available program such as, for example, the Wisconsin Package Version 10.0, such as BLAST, FASTA, PILEUP, FINDPATTERNS or the like (GCG, Madison, WI). Public sequence databases such as GenBank, EMBL, Swiss-Prot and PIR or private sequence databases such as PHYTOSEQ sequence database (Incyte Genomics, Palo Alto, CA) can be searched.

Alignment of sequences for comparison can be conducted by the local homology algorithm of Smith and Waterman (1981) Adv. Appl. Math. 2:482, by the homology alignment algorithm of Needleman and Wunsch (1970) J. Mol. Biol. 48:443-453, by the search for similarity method of Pearson and Lipman (1988) Proc. Natl. Acad. Sci. U.S.A. 85:2444-2448, by computerized implementations of these algorithms. After alignment, sequence comparisons between two (or more) polynucleotides or polypeptides are typically performed by comparing sequences of the two sequences over a comparison window to identify and compare local regions of sequence similarity. The comparison window can be a segment of at least about 20 contiguous positions, usually about 50 to about 200, more usually about 100 to about 150 contiguous positions. A description of the method is provided in Ausubel et al., *supra*.

A variety of methods for determining sequence relationships can be used, including manual alignment and computer assisted sequence alignment and analysis. This later approach is a preferred approach in the present invention, due to the increased throughput afforded by computer assisted methods. As noted above, a variety of computer programs for performing sequence alignment are available, or can be produced by one of skill.

One example algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul et al. J. Mol. Biol. 215:403-410 (1990). Software for performing BLAST analyses is publicly available, e.g., through the National Center for Biotechnology Information (see internet website at ncbi.nlm.nih.gov). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915). Unless otherwise indicated, "sequence identity" here refers to the % sequence identity generated from a tblastx using the NCBI version of

the algorithm at the default settings using gapped alignments with the filter "off" (see, for example, internet website at ncbi.nlm.nih.gov).

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.,* Karlin & Altschul (1993) Proc. Natl. Acad. Sci. USA 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability ($P(N)$), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence (and, therefore, in this context, homologous) if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, or less than about 0.01, and or even less than about 0.001. An additional example of a useful sequence alignment algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments. The program can align, e.g., up to 300 sequences of a maximum length of 5,000 letters.

The integrated system, or computer typically includes a user input interface allowing a user to selectively view one or more sequence records corresponding to the one or more character strings, as well as an instruction set which aligns the one or more character strings with each other or with an additional character string to identify one or more region of sequence similarity. The system may include a link of one or more character strings with a particular phenotype or gene function. Typically, the system includes a user readable output element that displays an alignment produced by the alignment instruction set.

The methods of this invention can be implemented in a localized or distributed computing environment. In a distributed environment, the methods may implemented on a single computer comprising multiple processors or on a multiplicity of computers. The computers can be linked, e.g. through a common bus, but more preferably the computer(s) are nodes on a network. The network can be a generalized or a dedicated local or wide-area network and, in certain preferred embodiments, the computers may be components of an intra-net or an internet.

Thus, the invention provides methods for identifying a sequence similar or homologous to one or more polynucleotides as noted herein, or one or more target polypeptides encoded by the polynucleotides, or otherwise noted herein and may include linking or associating a given plant phenotype or gene function with a sequence. In the methods, a sequence database is provided (locally or across an inter or intra net) and a query is made against the sequence database using the relevant sequences herein and associated plant phenotypes or gene functions.

Any sequence herein can be entered into the database, before or after querying the database. This provides for both expansion of the database and, if done before the querying step, for insertion of control sequences into the database. The control sequences can be detected by the query to ensure the general integrity of both the database and the query. As noted, the query can be performed using a web browser based interface. For example, the database can be a centralized public database such as those noted herein, and the querying can be done from a remote terminal or computer across an internet or intranet.

XVIII. Examples

The following examples are intended to illustrate but not limit the present invention. The complete descriptions of the traits associated with each polynucleotide of the invention is fully disclosed in Table 4 and Table 6.

Example I: Full Length Gene Identification and Cloning

Putative transcription factor sequences (genomic or ESTs) related to known transcription factors were identified in the *Arabidopsis thaliana* GenBank database using the tblastn sequence analysis program using default parameters and a P-value cutoff threshold of -4 or -5 or lower, depending on the length of the query sequence. Putative transcription factor sequence hits were then screened to identify those containing particular sequence strings. If the sequence hits contained such sequence strings, the sequences were confirmed as transcription factors.

Alternatively, *Arabidopsis thaliana* cDNA libraries derived from different tissues or treatments, or genomic libraries were screened to identify novel members of a transcription family using a low stringency hybridization approach. Probes were

synthesized using gene specific primers in a standard PCR reaction (annealing temperature 60° C) and labeled with ^{32}P dCTP using the High Prime DNA Labeling Kit (Boehringer Mannheim). Purified radiolabelled probes were added to filters immersed in Church hybridization medium (0.5 M NaPO_4 pH 7.0, 7% SDS, 1 % w/v bovine serum albumin) and hybridized overnight at 60°C with shaking. Filters were washed two times for 45 to 60 minutes with 1xSCC, 1% SDS at 60° C.

To identify additional sequence 5' or 3' of a partial cDNA sequence in a cDNA library, 5' and 3' rapid amplification of cDNA ends (RACE) was performed using the U.C. Marathon cDNA amplification kit (Clontech, Palo Alto, CA). Generally, the method entailed first isolating poly(A) mRNA, performing first and second strand cDNA synthesis to generate double stranded cDNA, blunting cDNA ends, followed by ligation of the U.C. Marathon Adaptor to the cDNA to form a library of adaptor-ligated ds cDNA.

Gene-specific primers were designed to be used along with adaptor specific primers for both 5' and 3' RACE reactions. Nested primers, rather than single primers, were used to increase PCR specificity. Using 5' and 3' RACE reactions, 5' and 3' RACE fragments were obtained, sequenced and cloned. The process can be repeated until 5' and 3' ends of the full-length gene were identified. Then the full-length cDNA was generated by PCR using primers specific to 5' and 3' ends of the gene by end-to-end PCR.

Example II: Construction of Expression Vectors

The sequence was amplified from a genomic or cDNA library using primers specific to sequences upstream and downstream of the coding region. The expression vector was pMEN20 or pMEN65, which are both derived from pMON316 (Sanders et al, (1987) Nucleic Acids Research 15:1543-1558) and contain the CaMV 35S promoter to express transgenes. To clone the sequence into the vector, both pMEN20 and the amplified DNA fragment were digested separately with SalI and NotI restriction enzymes at 37° C for 2 hours. The digestion products were subject to electrophoresis in a 0.8% agarose gel and visualized by ethidium bromide staining. The DNA fragments containing the sequence and the linearized plasmid were excised and purified by using a Qiaquick gel extraction kit (Qiagen, Valencia CA). The

fragments of interest were ligated at a ratio of 3:1 (vector to insert). Ligation reactions using T4 DNA ligase (New England Biolabs, Beverly MA) were carried out at 16° C for 16 hours. The ligated DNAs were transformed into competent cells of the *E. coli* strain DH5alpha by using the heat shock method. The transformations were plated on LB plates containing 50 mg/l kanamycin (Sigma, St. Louis, MO). Individual colonies were grown overnight in five milliliters of LB broth containing 50 mg/l kanamycin at 37° C. Plasmid DNA was purified by using Qiaquick Mini Prep kits (Qiagen).

Example III: Transformation of *Agrobacterium* with the Expression Vector

After the plasmid vector containing the gene was constructed, the vector was used to transform *Agrobacterium tumefaciens* cells expressing the gene products. The stock of *Agrobacterium tumefaciens* cells for transformation were made as described by Nagel et al. (1990) FEMS Microbiol Letts. 67: 325-328. *Agrobacterium* strain ABI was grown in 250 ml LB medium (Sigma) overnight at 28°C with shaking until an absorbance (A_{600}) of 0.5 – 1.0 was reached. Cells were harvested by centrifugation at 4,000 x g for 15 min at 4° C. Cells were then resuspended in 250 µl chilled buffer (1 mM HEPES, pH adjusted to 7.0 with KOH). Cells were centrifuged again as described above and resuspended in 125 µl chilled buffer. Cells were then centrifuged and resuspended two more times in the same HEPES buffer as described above at a volume of 100 µl and 750 µl, respectively. Resuspended cells were then distributed into 40 µl aliquots, quickly frozen in liquid nitrogen, and stored at -80° C.

Agrobacterium cells were transformed with plasmids prepared as described above following the protocol described by Nagel et al. For each DNA construct to be transformed, 50 – 100 ng DNA (generally resuspended in 10 mM Tris-HCl, 1 mM EDTA, pH 8.0) was mixed with 40 µl of *Agrobacterium* cells. The DNA/cell mixture was then transferred to a chilled cuvette with a 2mm electrode gap and subject to a 2.5 kV charge dissipated at 25 µF and 200 µF using a Gene Pulser II apparatus (Bio-Rad, Hercules, CA). After electroporation, cells were immediately resuspended in 1.0 ml LB and allowed to recover without antibiotic selection for 2 – 4 hours at 28° C in a shaking incubator. After recovery, cells were plated onto selective medium of LB broth containing 100 µg/ml spectinomycin (Sigma) and incubated for 24-48 hours at 28° C. Single colonies were then picked and inoculated in fresh medium. The

presence of the plasmid construct was verified by PCR amplification and sequence analysis.

Example IV: Transformation of *Arabidopsis* Plants with *Agrobacterium tumefaciens* with Expression Vector

After transformation of *Agrobacterium tumefaciens* with plasmid vectors containing the gene, single *Agrobacterium* colonies were identified, propagated, and used to transform *Arabidopsis* plants. Briefly, 500 ml cultures of LB medium containing 50 mg/l kanamycin were inoculated with the colonies and grown at 28° C with shaking for 2 days until an optical absorbance at 600 nm wavelength over 1 cm (A_{600}) of > 2.0 is reached. Cells were then harvested by centrifugation at 4,000 x g for 10 min, and resuspended in infiltration medium (1/2 X Murashige and Skoog salts (Sigma), 1 X Gamborg's B-5 vitamins (Sigma), 5.0% (w/v) sucrose (Sigma), 0.044 μ M benzylamino purine (Sigma), 200 μ l/l Silwet L-77 (Lehle Seeds) until an A_{600} of 0.8 was reached.

Prior to transformation, *Arabidopsis thaliana* seeds (ecotype Columbia) were sown at a density of ~10 plants per 4" pot onto Pro-Mix BX potting medium (Hummert International) covered with fiberglass mesh (18 mm X 16 mm). Plants were grown under continuous illumination (50-75 μ E/m²/sec) at 22-23° C with 65-70% relative humidity. After about 4 weeks, primary inflorescence stems (bolts) are cut off to encourage growth of multiple secondary bolts. After flowering of the mature secondary bolts, plants were prepared for transformation by removal of all siliques and opened flowers.

The pots were then immersed upside down in the mixture of *Agrobacterium* infiltration medium as described above for 30 sec, and placed on their sides to allow draining into a 1' x 2' flat surface covered with plastic wrap. After 24 h, the plastic wrap was removed and pots are turned upright. The immersion procedure was repeated one week later, for a total of two immersions per pot. Seeds were then collected from each transformation pot and analyzed following the protocol described below.

Example V: Identification of *Arabidopsis* Primary Transformants

Seeds collected from the transformation pots were sterilized essentially as follows. Seeds were dispersed into in a solution containing 0.1% (v/v) Triton X-100 (Sigma) and sterile H₂O and washed by shaking the suspension for 20 min. The wash solution was then drained and replaced with fresh wash solution to wash the seeds for 20 min with shaking. After removal of the second wash solution, a solution containing 0.1% (v/v) Triton X-100 and 70% ethanol (Equistar) was added to the seeds and the suspension was shaken for 5 min. After removal of the ethanol/detergent solution, a solution containing 0.1% (v/v) Triton X-100 and 30% (v/v) bleach (Clorox) was added to the seeds, and the suspension was shaken for 10 min. After removal of the bleach/detergent solution, seeds were then washed five times in sterile distilled H₂O. The seeds were stored in the last wash water at 4° C for 2 days in the dark before being plated onto antibiotic selection medium (1 X Murashige and Skoog salts (pH adjusted to 5.7 with 1M KOH), 1 X Gamborg's B-5 vitamins, 0.9% phytagar (Life Technologies), and 50 mg/l kanamycin). Seeds were germinated under continuous illumination (50-75 $\mu\text{E}/\text{m}^2/\text{sec}$) at 22-23° C. After 7-10 days of growth under these conditions, kanamycin resistant primary transformants (T₁ generation) were visible and obtained. These seedlings were transferred first to fresh selection plates where the seedlings continued to grow for 3-5 more days, and then to soil (Pro-Mix BX potting medium).

Primary transformants were crossed and progeny seeds (T₂) collected; kanamycin resistant seedlings were selected and analyzed. The expression levels of the recombinant polynucleotides in the transformants varies from about a 5% expression level increase to a least a 100% expression level increase. Similar observations are made with respect to polypeptide level expression.

Example VI: Identification of *Arabidopsis* Plants with Transcription Factor Gene Knockouts

The screening of insertion mutagenized *Arabidopsis* collections for null mutants in a known target gene was essentially as described in Krysan et al (1999) Plant Cell 11:2283-2290. Briefly, gene-specific primers, nested by 5-250 base pairs to each other, were designed from the 5' and 3' regions of a known target gene. Similarly, nested sets of primers were also created specific to each of the T-DNA or

transposon ends (the "right" and "left" borders). All possible combinations of gene specific and T-DNA/transposon primers were used to detect by PCR an insertion event within or close to the target gene. The amplified DNA fragments were then sequenced which allows the precise determination of the T-DNA/transposon insertion point relative to the target gene. Insertion events within the coding or intervening sequence of the genes were deconvoluted from a pool comprising a plurality of insertion events to a single unique mutant plant for functional characterization. The method is described in more detail in Yu and Adam, US Application Serial No. 09/177,733 filed October 23, 1998.

Example VII: Identification of Modified Phenotypes in Overexpression or Gene Knockout Plants

Experiments were performed to identify those transformants or knockouts that exhibited modified biochemical characteristics. Among the biochemicals that were assayed were insoluble sugars, such as arabinose, fucose, galactose, mannose, rhamnose or xylose or the like; prenyl lipids, such as lutein, beta-carotene, xanthophyll-1, xanthophyll-2, chlorophylls A or B, or alpha-, delta- or gamma-tocopherol or the like; fatty acids, such as 16:0 (palmitic acid), 16:1 (palmitoleic acid), 18:0 (stearic acid), 18:1 (oleic acid), 18:2 (linoleic acid), 20:0, 18:3 (linolenic acid), 20:1 (eicosenoic acid), 20:2, 22:1 (erucic acid) or the like; waxes, such as by altering the levels of C29, C31, or C33 alkanes; sterols, such as brassicasterol, campesterol, stigmasterol, sitosterol or stigmastanol or the like, glucosinolates, protein or oil levels.

Fatty acids were measured using two methods depending on whether the tissue was from leaves or seeds. For leaves, lipids were extracted and esterified with hot methanolic H_2SO_4 and partitioned into hexane from methanolic brine. For seed fatty acids, seeds were pulverized and extracted in methanol:heptane:toluene:2,2-dimethoxypropane: H_2SO_4 (39:34:20:5:2) for 90 minutes at 80°C. After cooling to room temperature the upper phase, containing the seed fatty acid esters, was subjected to GC analysis. Fatty acid esters from both seed and leaf tissues were analyzed with a Supelco SP-2330 column.

Glucosinolates were purified from seeds or leaves by first heating the tissue at 95°C for 10 minutes. Preheated ethanol:water (50:50) is and after heating at 95°C for a further 10 minutes, the extraction solvent is applied to a DEAE Sephadex column which had been previously equilibrated with 0.5 M pyridine acetate. Desulfoglucosinolates were eluted with 300 ul water and analyzed by reverse phase HPLC monitoring at 226 nm.

For wax alkanes, samples were extracted using an identical method as fatty acids and extracts were analyzed on a HP 5890 GC coupled with a 5973 MSD. Samples were chromatographically isolated on a J&W DB35 mass spectrometer (J&W Scientific).

To measure prenyl lipids levels, seeds or leaves were pulverized with 1 to 2% pyrogallol as an antioxidant. For seeds, extracted samples were filtered and a portion removed for tocopherol and carotenoid/chlorophyll analysis by HPLC. The remaining material was saponified for sterol determination. For leaves, an aliquot was removed and diluted with methanol and chlorophyll A, chlorophyll B, and total carotenoids measured by spectrophotometry by determining optical absorbance at 665.2 nm, 652.5 nm, and 470 nm. An aliquot was removed for tocopherol and carotenoid/chlorophyll composition by HPLC using a Waters uBondapak C18 column (4.6 mm x 150 mm). The remaining methanolic solution was saponified with 10% KOH at 80°C for one hour. The samples were cooled and diluted with a mixture of methanol and water. A solution of 2% methylene chloride in hexane was mixed in and the samples were centrifuged. The aqueous methanol phase was again re-extracted 2% methylene chloride in hexane and, after centrifugation, the two upper phases were combined and evaporated. 2% methylene chloride in hexane was added to the tubes and the samples were then extracted with one ml of water. The upper phase was removed, dried, and resuspended in 400 ul of 2% methylene chloride in hexane and analyzed by gas chromatography using a 50 m DB-5ms (0.25 mm ID, 0.25 um phase, J&W Scientific).

Insoluble sugar levels were measured by the method essentially described by Reiter et al., (1997) Plant Journal 12:335-345. This method analyzes the neutral sugar composition of cell wall polymers found in *Arabidopsis* leaves. Soluble sugars were

separated from sugar polymers by extracting leaves with hot 70% ethanol. The remaining residue containing the insoluble polysaccharides was then acid hydrolyzed with allose added as an internal standard. Sugar monomers generated by the hydrolysis were then reduced to the corresponding alditols by treatment with NaBH₄, then were acetylated to generate the volatile alditol acetates which were then analyzed by GC-FID. Identity of the peaks was determined by comparing the retention times of known sugars converted to the corresponding alditol acetates with the retention times of peaks from wild-type plant extracts. Alditol acetates were analyzed on a Supelco SP-2330 capillary column (30 m x 250 μ m x 0.2 μ m) using a temperature program beginning at 180° C for 2 minutes followed by an increase to 220° C in 4 minutes. After holding at 220° C for 10 minutes, the oven temperature is increased to 240° C in 2 minutes and held at this temperature for 10 minutes and brought back to room temperature.

To identify plants with alterations in total seed oil or protein content, 150mg of seeds from T2 progeny plants were subjected to analysis by Near Infrared Reflectance Spectroscopy (NIRS) using a Foss NirSystems Model 6500 with a spinning cup transport system. NIRS is a non-destructive analytical method used to determine seed oil and protein composition. Infrared is the region of the electromagnetic spectrum located after the visible region in the direction of longer wavelengths. 'Near infrared' owns its name for being the infrared region near to the visible region of the electromagnetic spectrum. For practical purposes, near infrared comprises wavelengths between 800 and 2500 nm. NIRS is applied to organic compounds rich in O-H bonds (such as moisture, carbohydrates, and fats), C-H bonds (such as organic compounds and petroleum derivatives), and N-H bonds (such as proteins and amino acids). The NIRS analytical instruments operate by statistically correlating NIRS signals at several wavelengths with the characteristic or property intended to be measured. All biological substances contain thousands of C-H, O-H, and N-H bonds. Therefore, the exposure to near infrared radiation of a biological sample, such as a seed, results in a complex spectrum which contains qualitative and quantitative information about the physical and chemical composition of that sample.

The numerical value of a specific analyte in the sample, such as protein content or oil content, is mediated by a calibration approach known as chemometrics.

Chemometrics applies statistical methods such as multiple linear regression (MLR), partial least squares (PLS), and principle component analysis (PCA) to the spectral data and correlates them with a physical property or other factor, that property or factor is directly determined rather than the analyte concentration itself. The method first provides "wet chemistry" data of the samples required to develop the calibration.

Calibration for Arabidopsis seed oil composition was performed using accelerated solvent extraction using 1 g seed sample size and was validated against certified canola seed. A similar wet chemistry approach was performed for seed protein composition calibration.

Data obtained from NIRS analysis was analyzed statistically using a nearest-neighbor (N-N) analysis. The N-N analysis allows removal of within-block spatial variability in a fairly flexible fashion which does not require prior knowledge of the pattern of variability in the chamber. Ideally, all hybrids are grown under identical experimental conditions within a block (rep). In reality, even in many block designs, significant within-block variability exists. Nearest-neighbor procedures are based on assumption that environmental effect of a plot is closely related to that of its neighbors. Nearest-neighbor methods use information from adjacent plots to adjust for within-block heterogeneity and so provide more precise estimates of treatment means and differences. If there is within-plot heterogeneity on a spatial scale that is larger than a single plot and smaller than the entire block, then yields from adjacent plots will be positively correlated. Information from neighboring plots can be used to reduce or remove the unwanted effect of the spatial heterogeneity, and hence improve the estimate of the treatment effect. Data from neighboring plots can also be used to reduce the influence of competition between adjacent plots. The Papadakis N-N analysis can be used with designs to remove within-block variability that would not be removed with the standard split plot analysis (Papadakis, 1973, Inst. d'Amelior. Plantes Thessaloniki (Greece) Bull. Scientif., No. 23; Papadakis, 1984, Proc. Acad. Athens, 59, 326-342).

Experiments were performed to identify those transformants or knockouts that exhibited an improved pathogen tolerance. For such studies, the transformants were exposed to biotrophic fungal pathogens, such as *Erysiphe orontii*, and necrotropic

fungal pathogens, such as *Fusarium oxysporum*. *Fusarium oxysporum* isolates cause vascular wilts and damping off of various annual vegetables, perennials and weeds (Mauch-Mani and Slusarenko (1994) Molecular Plant-Microbe Interactions 7: 378-383). For *Fusarium oxysporum* experiments, plants grown on Petri dishes were sprayed with a fresh spore suspension of *F. oxysporum*. The spore suspension was prepared as follows: A plug of fungal hyphae from a plate culture was placed on a fresh potato dextrose agar plate and allowed to spread for one week. 5 ml sterile water was then added to the plate, swirled, and pipetted into 50 ml Armstrong *Fusarium* medium. Spores were grown overnight in *Fusarium* medium and then sprayed onto plants using a Preval paint sprayer. Plant tissue was harvested and frozen in liquid nitrogen 48 hours post infection.

Erysiphe orontii is a causal agent of powdery mildew. For *Erysiphe orontii* experiments, plants were grown approximately 4 weeks in a greenhouse under 12 hour light (20°C, ~30% relative humidity (rh)). Individual leaves were infected with *E. orontii* spores from infected plants using a camel's hair brush, and the plants were transferred to a Percival growth chamber (20°C, 80% rh.). Plant tissue was harvested and frozen in liquid nitrogen 7 days post infection.

Botrytis cinerea is a necrotrophic pathogen. *Botrytis cinerea* was grown on potato dextrose agar in the light. A spore culture was made by spreading 10 ml of sterile water on the fungus plate, swirling and transferring spores to 10 ml of sterile water. The spore inoculum (approx. 105 spores/ml) was used to spray 10 day-old seedlings grown under sterile conditions on MS (minus sucrose) media. Symptoms were evaluated every day up to approximately 1 week.

Infection with bacterial pathogens *Pseudomonas syringae* pv *maculicola* (Psm) strain 4326 and pv *maculicola* strain 4326 was performed by hand inoculation at two doses. Two inoculation doses allows the differentiation between plants with enhanced susceptibility and plants with enhanced resistance to the pathogen. Plants were grown for 3 weeks in the greenhouse, then transferred to the growth chamber for the remainder of their growth. Psm ES4326 was hand inoculated with 1 ml syringe on 3 fully-expanded leaves per plant (4 1/2 wk old), using at least 9 plants per overexpressing line at two inoculation doses, OD=0.005 and OD=0.0005. Disease

scoring occurred at day 3 post-inoculation with pictures of the plants and leaves taken in parallel.

In some instances, expression patterns of the pathogen-induced genes (such as defense genes) was monitored by microarray experiments. cDNAs were generated by PCR and resuspended at a final concentration of ~ 100 ng/ul in 3X SSC or 150mM Na-phosphate (Eisen and Brown (1999) *Methods Enzymol.* 303:179-205). The cDNAs were spotted on microscope glass slides coated with polylysine. The prepared cDNAs were aliquoted into 384 well plates and spotted on the slides using an x-y-z gantry (OmniGrid) purchased from GeneMachines (Menlo Park, CA) outfitted with quill type pins purchased from Telechem International (Sunnyvale, CA). After spotting, the arrays were cured for a minimum of one week at room temperature, rehydrated and blocked following the protocol recommended by Eisen and Brown (1999; *supra*).

Sample total RNA (10 ug) samples were labeled using fluorescent Cy3 and Cy5 dyes. Labeled samples were resuspended in 4X SSC/0.03% SDS/4 ug salmon sperm DNA/2 ug tRNA/ 50mM Na-pyrophosphate, heated for 95°C for 2.5 minutes, spun down and placed on the array. The array was then covered with a glass coverslip and placed in a sealed chamber. The chamber was then kept in a water bath at 62°C overnight. The arrays were washed as described in Eisen and Brown (1999) and scanned on a General Scanning 3000 laser scanner. The resulting files are subsequently quantified using Imogene, a software purchased from BioDiscovery (Los Angeles, CA).

Experiments were performed to identify those transformants or knockouts that exhibited an improved environmental stress tolerance. For such studies, the transformants were exposed to a variety of environmental stresses. Plants were exposed to chilling stress (6 hour exposure to 4-8°C), heat stress (6 hour exposure to 32-37°C), high salt stress (6 hour exposure to 200 mM NaCl), drought stress (168 hours after removing water from trays), osmotic stress (6 hour exposure to 3 M mannitol), or nutrient limitation (nitrogen, phosphate, and potassium) (Nitrogen: all components of MS medium remained constant except N was reduced to 20 mg/l of NH₄NO₃, or Phosphate: All components of MS medium except KH₂PO₄, which was

replaced by K_2SO_4 , Potassium: All components of MS medium except removal of KNO_3 and KH_2PO_4 , which were replaced by NaH_4PO_4).

Experiments were performed to identify those transformants or knockouts that exhibited a modified structure and development characteristics. For such studies, the transformants were observed by eye to identify novel structural or developmental characteristics associated with the ectopic expression of the polynucleotides or polypeptides of the invention.

Experiments were performed to identify those transformants or knockouts that exhibited modified sugar-sensing. For such studies, seeds from transformants were germinated on media containing 5% glucose or 9.4% sucrose which normally partially restrict hypocotyl elongation. Plants with altered sugar sensing may have either longer or shorter hypocotyls than normal plants when grown on this media. Additionally, other plant traits may be varied such as root mass.

Flowering time was measured by the number of rosette leaves present when a visible inflorescence of approximately 3 cm is apparent. Rosette and total leaf number on the progeny stem are tightly correlated with the timing of flowering (Koornneef et al (1991) *Mol. Gen. Genet* 229:57-66. The vernalization response was measured. For vernalization treatments, seeds were sown to MS agar plates, sealed with micropore tape, and placed in a 4°C cold room with low light levels for 6-8 weeks. The plates were then transferred to the growth rooms alongside plates containing freshly sown non-vernalized controls. Rosette leaves were counted when a visible inflorescence of approximately 3 cm was apparent.

Modified phenotypes observed for particular overexpressor or knockout plants are provided in Table 4. For a particular overexpressor that shows a less beneficial characteristic, it may be more useful to select a plant with a decreased expression of the particular transcription factor. For a particular knockout that shows a less beneficial characteristic, it may be more useful to select a plant with an increased expression of the particular transcription factor.

The sequences of the Sequence Listing or those in Tables 4, 5 or those disclosed here can be used to prepare transgenic plants and plants with altered traits. The specific transgenic plants listed below are produced from the sequences of the Sequence Listing, as noted. Table 4 provides exemplary polynucleotide and polypeptide sequences of the invention. Table 4 includes, from left to right for each sequence: the first column shows the polynucleotide SEQ ID NO; the second column shows the Mendel Gene ID No., GID; the third column shows the trait(s) resulting from the knock out or overexpression of the polynucleotide in the transgenic plant; the fourth column shows the category of the trait; the fifth column shows the transcription factor family to which the polynucleotide belongs; the sixth column ("Comment"), includes specific effects and utilities conferred by the polynucleotide of the first column; the seventh column shows the SEQ ID NO of the polypeptide encoded by the polynucleotide; and the eighth column shows the amino acid residue positions of the conserved domain in amino acid (AA) co-ordinates.

Seed of plants overexpressing sequences G265 (SEQ ID NOs:871 and 872), G715 (SEQ ID NOs:925 and 926), G1471 (SEQ ID NOs:311 and 312), G1793 (SEQ ID NOs:365 and 366), G1838 (SEQ ID NOs:381 and 382), G1902 (SEQ ID NOs:405 and 406), G286 (SEQ ID NOs:877 and 878), G2138 (SEQ ID NOs:865 and 866) and G2830 (SEQ ID NOs:875 and 876) was subjected to NIR analysis and a significant increase in seed oil content compared with seed from control plants was identified.

G192: G192 (SEQ ID NO: 859) was expressed in all plant tissues and under all conditions examined. Its expression was slightly induced upon infection by *Fusarium*. G192 was analyzed using transgenic plants in which this gene was expressed under the control of the 35S promoter. G192 overexpressors were late flowering under 12 hour light and had more leaves than control plants. This phenotype was manifested in the three T2 lines analyzed. Results of one experiment suggest that G192 overexpressor was more susceptible to infection with a moderate dose of the fungal pathogen *Erysiphe orontii*. The decrease in seed oil observed for one line was replicated in an independent experiment. G192 overexpression delayed flowering. A wide variety of applications exist for systems that either lengthen or shorten the time to flowering, or for systems of inducible flowering time control. In particular, in species where the vegetative parts of the plants constitute the crop and

the reproductive tissues are discarded, it will be advantageous to delay or prevent flowering. Extending vegetative development can bring about large increases in yields. G192 can be used to manipulate the defense response in order to generate pathogen-resistant plants. G192 can be used to manipulate seed oil content, which can be of nutritional value.

Closely Related Genes from Other Species

G192 had some similarity within the conserved WRKY domain to non-Arabidopsis plant proteins.

G1946: G1946 (SEQ ID NO: 801) was studied using transgenic plants in which the gene was expressed under the control of the 35S promoter. Overexpression of G1946 resulted in accelerated flowering, with 35S::G1946 transformants producing flower buds up to a week earlier than wild-type controls (24-hour light conditions). These effects were seen in 12/20 primary transformants and in two independent plantings of each of the three T2 lines. Unlike many early flowering Arabidopsis transgenic lines, which are dwarfed, 35S::G1946 transformants often reached full-size at maturity, and produced large quantities of seeds, although the plants were slightly pale in coloration and had slightly flat leaves compared to wild-type. In addition, 35S::G1946 plants showed an altered response to phosphate deprivation. Seedlings of G1946 overexpressor plants showed more secondary root growth on phosphate-free media, when compared to wild-type control. In a repeat experiment, all three lines showed the phenotype. Overexpression of G1946 in Arabidopsis also resulted in an increase in seed glucosinolate M39501 in T2 lines 1 and 3. An increase in seed oil and a decrease in seed protein was also observed in these two lines. G1946 was ubiquitously expressed, and does not appear to be significantly induced or repressed by any of the biotic and abiotic stress conditions tested at this time, with the exception of cold, which repressed G1946 expression. G1946 can be used to modify flowering time, as well as to improve the plant's performance in conditions of limited phosphate, and to alter seed oil, protein, and glucosinolate composition.

Closely Related Genes from Other Species

A comparison of the amino acid sequence of G1946 with sequences available from GenBank showed strong similarity with plant HSFs of several species (*Lycopersicon peruvianum*, *Medicago truncatula*, *Lycopersicon esculentum*, *Glycine max*, *Solanum tuberosum*, *Oryza sativa* and *Hordeum vulgare* subsp. *vulgare*).

G375: The sequence of G375 (SEQ ID NO:239) was experimentally determined and G375 was analyzed using transgenic plants in which G375 was expressed under the control of the 35S promoter. Overexpression of G375 produced marked effects on leaf development. At early stages of growth, 35S::G375 seedlings developed narrow, upward pointing leaves with long petioles (possibly indicating a disruption in circadian-clock controlled processes or nyctinastic movements). Additionally, some seedlings were noted to have elongated hypocotyls, and some were rather small compared to wild-type controls. Comparable phenotypes were obtained by overexpression of an AP2 family gene, G2113 (SEQ ID NO: 85). Following the switch to flowering, 35S::G375 plants showed reduced fertility, which possibly arose from a failure of stamens to fully elongate. One of the three T2 lines, (#41) was later flowering than wild-type controls, and also developed large numbers of small secondary rosette leaves in the axils of the primary rosette. Although these effects were not noted in the other two lines, the phenotypes obtained in line 41 were somewhat similar to those produced by overexpression of another Z-dof gene, G736 (SEQ ID NO: 211). G375 was expressed in all tissues, although at different levels. It was expressed at low levels in the root and germinating seed, and expressed at high levels in the embryo. The effects of G375 on leaf architecture are of potential interest to the ornamental horticulture industry.

Closely Related Genes from Other Species

G375 showed some homology to non-Arabidopsis plant proteins within the conserved Dof domain.

G1255: The sequence of G1255 (SEQ ID NO: 273) was experimentally determined and G1255 was analyzed using transgenic plants in which G1255 was expressed under the control of the 35S promoter. Plants overexpressing G1255 had

alterations in leaf architecture, a reduction in apical dominance, an increase in seed size, and showed more disease symptoms following inoculation with a low dose of the fungal pathogen *Botrytis cinerea*. G1255 was constitutively expressed and not significantly induced by any conditions tested. On the basis of the phenotypes produced by overexpression of G1255, G1255 can be used to manipulate the plant's defense response to produce pathogen resistance, alter plant architecture, or alter seed size.

Closely Related Genes from Other Species

G1255 showed strong homology to a putative rice zing finger protein represented by sequence AC087181_3. Sequence identity between these two protein extended beyond the conserved domain, and therefore, these genes can be orthologs.

G865: The complete cDNA sequence of G865 (SEQ ID NO: 557) was determined. G865 was ubiquitously expressed in Arabidopsis tissues. G865 was analyzed using transgenic plants in which G865 was expressed under the control of the 35S promoter. Plants overexpressing G865 were early flowering, with numerous secondary inflorescence meristems giving them a bushy appearance. G865 overexpressors were more susceptible to infection with a moderate dose of the fungal pathogens *Erysiphe orontii* and *Botrytis cinerea*. In addition, seeds from G865 overexpressing plants showed a trend of increased protein and reduced oil content, although the observed changes were not beyond the criteria used for judging significance except in one line. G865 can be used to control flowering time. G865 can be used to manipulate the defense response in order to generate pathogen-resistant plants. G865 can be used to alter seed oil and protein content of a plant.

Closely Related Genes from Other Species

G865 and other non-Arabidopsis AP2/EREBP proteins were similar within the conserved AP2 domain.

G2509: G2509 (SEQ ID NO: 23) was studied using transgenic plants in which the gene was expressed under the control of the 35S promoter. Overexpression of G2509 caused multiple alterations in plant growth and development, most notably, altered branching patterns, and a reduction in apical dominance, giving the plants a

shorter, more bushy stature than wild type. Twenty 35S::G2509 primary transformants were examined; at early stages of rosette development, these plants displayed a wild-type phenotype. However, at the switch to flowering, almost all T1 lines showed a marked loss of apical dominance and large numbers of secondary shoots developed from axils of primary rosette leaves. In the most extreme cases, the shoots had very short internodes, giving the inflorescence a very bushy appearance. Such shoots were often very thin and flowers were relatively small and poorly fertile. At later stages, many plants appeared very small and had a low seed yield compared to wild type. In addition to the effects on branching, a substantial number of 35S::G2509 primary transformants also flowered early and had buds visible several days prior to wild type. Similar effects on inflorescence development were noted in each of three T2 populations examined. The branching and plant architecture phenotypes observed in 35S::G2509 lines resemble phenotypes observed for three other AP2/EREBP genes: G865 (SEQ ID NO: 557), G1411 (SEQ ID NO: 3), and G1794 (SEQ ID NO: 13). G2509, G865, and G1411 form a small clade within the large AP2/EREBP family, and G1794, although not belonging to the clade, is one of the AP2/EREBP genes closest to it in the phylogenetic tree. It is thus likely that all these genes share a related function, such as affecting hormone balance. Overexpression of G2509 in Arabidopsis resulted in an increase in alpha-tocopherol in seeds in T2 lines 5 and 11. G2509 was ubiquitously expressed in Arabidopsis plant tissue. G2509 expression levels were altered by a variety of environmental or physiological conditions. G2509 can be used to manipulate plant architecture and development. G2509 can be used to alter tocopherol composition. Tocopherols have anti-oxidant and vitamin E activity. G2509 can be useful in altering flowering time. A wide variety of applications exist for systems that either lengthen or shorten the time to flowering.

Closely Related Genes from Other Species

G2509 showed some sequence similarity with known genes from other plant species within the conserved AP2/EREBP domain.

G2347: G2347 (SEQ ID NO: 1119) was analyzed using transgenic plants in which G2347 was expressed under the control of the 35S promoter. Overexpression of G2347 markedly reduced the time to flowering in Arabidopsis. This phenotype

was apparent in the majority of primary transformants and in all plants from two out of the three T2 lines examined. Under continuous light conditions, 35S::G2347 plants formed flower buds up a week earlier than wild type. Many of the plants were rather small and spindly compared to controls. To demonstrate that overexpression of G2347 could induce flowering under less inductive photoperiods, two T2 lines were re-grown in 12 hour conditions; again, all plants from both lines bolted early, with some initiating flower buds up to two weeks sooner than wild-type. As determined by RT-PCR, G2347 was highly expressed in rosette leaves and flowers, and to much lower levels in embryos and siliques. No expression of G2347 was detected in the other tissues tested. G2347 expression was repressed by cold, and by auxin treatments and by infection by Erysiphe. G2347 is also highly similar to the Arabidopsis protein G2010 (SEQ ID NO: 1121). The level of homology between these two proteins suggested they could have similar, overlapping, or redundant functions in Arabidopsis. In support of this hypothesis, overexpression of both G2010 and G2347 resulted in early flowering phenotypes in transgenic plants.

Closely Related Genes from Other Species

The closest relative to G2347 is the Antirrhinum protein, SBP2 (CAA63061). The similarity between these two proteins is extensive enough to suggest they might have similar functions in a plant.

G988: G988 (SEQ ID NO: 43) was analyzed using transgenic plants in which G988 was expressed under the control of the 35S promoter. Plants overexpressing G988 had multiple morphological phenotypes. The transgenic plants were generally smaller than wild-type plants, had altered leaf, inflorescence and flower development, altered plant architecture, and altered vasculature. In one transgenic line overexpressing G988 (line 23), an increase in the seed glucosinolate M39489 was observed. The phenotype of plants overexpressing G988 was wild-type in all other assays performed. In wild-type plants, G988 was expressed primarily in flower and silique tissue, but was also present at detectable levels in all other tissues tested. Expression of G988 was induced in response to heat treatment, and repressed in response to infection with Erysiphe. Based on the observed morphological phenotypes of the transgenic plants, G988 can be used to create plants with larger flowers. This can have value in the ornamental horticulture industry. The reduction

in the formation of lateral branches suggests that G988 can have utility on the forestry industry. The Arabidopsis plants overexpressing G988 also had reduced fertility. This can be a desirable trait in some instances, as it can be exploited to prevent or minimize the escape of GMO (genetically modified organism) pollen into the environment.

Closely Related Genes from Other Species

The amino acid sequence for the Capsella rubella hypothetical protein represented by GenBank accession number CRU303349 was significantly identical to G988 outside of the SCR conserved domains. The Capsella rubella hypothetical protein is 90% identical to G988 over a stretch of roughly 450 amino acids. Therefore, it is likely that the Capsella rubella gene is an ortholog of G988.

G2346: G2346 (SEQ ID NO: 459) was analyzed using transgenic plants in which the gene was expressed under the control of the 35S promoter. 35S::G2346 seedlings from all three T2 populations had slightly larger cotyledons and appeared somewhat more advanced than controls. This indicated that the seedlings developed more rapidly than the control plants. At later stages, however, G2346 overexpressing plants showed no consistent differences from control plants. The phenotype of these transgenic plants was wild-type in all other assays performed. According to RT-PCR analysis, G2346 is expressed ubiquitously.

Closely Related Genes from Other Species

G2346 shows some sequence similarity with known genes from other plant species within the conserved SBP domain.

G1354: The complete sequence of G1354 (SEQ ID NO: 285) was determined. G1354 was analyzed using transgenic plants in which G1354 was expressed under the control of the 35S promoter. Overexpression of G1354 produced highly deleterious effects on growth and development. Only three 35S::G1354 T1 plants were obtained; all were extremely tiny and slow developing. After three weeks of growth, each of the plants comprised a completely disorganized mass of leaves and root that had no clear axis of growth. Since these individuals would not have survived transplantation to soil, they were harvested for RT-PCR analysis; all three plants showed moderate

levels of G1354 overexpression compared to whole wild-type seedlings of an equivalent size. Only a very small number of transformants were obtained from two selection attempts on separate batches of T0 seed. Usually between 15 and 120 transformants are obtained from each aliquot of 300 mg T0 seed from wild-type plants. The low transformation frequency obtained in this experiment suggests that high levels of G1354 overexpression might have completely lethal effects and prevent transformed seeds from germinating. As determined by RT-PCR, G1354 was uniformly expressed in all tissues and under all conditions tested in RT-PCR. However, the gene was repressed in leaf tissue in response to Erysiphe infection.

Closely Related Genes from Other Species

G1354 is closely related to a NAM protein encoded by polynucleotide from rice (AC005310). Similarity between G1354 and this rice protein extends beyond the signature motif of the family to a level that would suggest the genes are orthologs.

G1063: G1063 (SEQ ID NO: 119) is a member of a clade of highly related HLH/MYC proteins that also includes G779 (SEQ ID NO: 113), G1499 (SEQ ID NO: 7), G2143 (SEQ ID NO: 129), and G2557 (SEQ ID NO: 133). All of these genes caused similar pleiotropic phenotypic effects when overexpressed, the most striking of which was the production of ectopic carpelloid tissue. These genes can be considered key regulators of carpel development. A spectrum of developmental alterations was observed amongst 35S::G1063 primary transformants and the majority were markedly small, dark green, and had narrow curled leaves. The most severely affected individuals were completely sterile and formed highly abnormal inflorescences; shoots often terminated in pin-like structures, and flowers were replaced by filamentous carpelloid structures. In other cases, flowers showed internode elongation between floral whorls, with a central carpel protruding on a pedicel-like organ. Additionally, lateral branches sometimes failed to develop and tiny patches of carpelloid tissue formed at axillary nodes of the inflorescence. In lines with an intermediate phenotype, flowers contained defined whorls of organs, but sepals were converted to carpelloid structures or displayed patches of carpelloid tissue. In contrast, lines with a weak phenotype developed relatively normal flowers and produced a reasonable quantity of seed. Such plants were still distinctly smaller than wild-type controls. Since the strongest 35S::G1063 lines were sterile, three lines

with a relatively weak phenotype, that had produced sufficient seed for biochemical and physiological analysis, were selected for further study. Two of the T2 populations (T2-28,37) were clearly small, darker green and possessed narrow leaves compared to wild type. Plants from one of these populations (T2-28) also produced occasional branches with abnormal flowers like those seen in the T1. The final T2 population (T2-30) displayed a very mild phenotype. Overexpression of G1063 in *Arabidopsis* resulted in a decrease in seed oil content in T2 lines 28 and 37. No altered phenotypes were detected in any of the physiological assays, except that the plants were noted to be somewhat small and produce anthocyanin when grown in Petri plates. G1063 was expressed at low to moderate levels in roots, flowers, rosette leaves, embryos, and germinating seeds, but was not detected in shoots or siliques. It was induced by auxin. G1063 can be used to manipulate flower form and structure or plant fertility. One application for manipulation of flower structure can be in the production of saffron, which is derived from the stigmas of *Crocus sativus*. G1063 has utility in manipulating seed oil and protein content.

Closely Related Genes from Other Species

G1063 protein shared extensive homology in the basic helix loop helix region with a protein sequence encoded by Glycine max cDNA clone (AW832545) as well as a tomato root, plants pre-anthesis *Lycopersicon sculentum* cDNA (BE451174).

G2143: G2143 (SEQ ID NO: 129) is a member of a clade of highly related HLH/MYC proteins that also includes G779 (SEQ ID NO: 113), G1063 (SEQ ID NO: 119), G1499 (SEQ ID NO: 7), and G2557 (SEQ ID NO: 133). All of these genes caused similar pleiotropic phenotypic effects when overexpressed, the most striking of which was the production of ectopic carpelloid tissue. These genes can be considered key regulators of carpel development. Twelve out of twenty 35S::G2143 T1 lines showed a very severe phenotype; these plants were markedly small and had narrow, curled, dark-green leaves. Such individuals were completely sterile and formed highly abnormal inflorescences; shoots often terminated in pin-like structures, and flowers were replaced by filamentous carpelloid structures, or a fused mass of carpelloid tissue. Furthermore, lateral branches usually failed to develop, and tiny patches of stigmatic tissue often formed at axillary nodes of the inflorescence. Strongly affected plants displayed the highest levels of transgene expression

(determined by RT-PCR). The remaining T1 lines showed lower levels of G2143 overexpression; these plants were still distinctly smaller than wild type, but had relatively normal inflorescences and produced seed. Since the strongest 35S::G2143 lines were sterile, three lines with a relatively weak phenotype, that had produced sufficient seed for biochemical analysis, were selected for further study. T2-11 plants displayed a very mild phenotype and had somewhat small, narrow, dark green leaves. The other two T2 populations, however, appeared wild-type, suggesting that transgene activity might have been reduced between the generations. Reduced seedling vigor was noted in the physiological assays. G2143 expression was detected at low levels in flowers and siliques, and at higher levels in germinating seed. G2143 can be used to manipulate flower form and structure or plant fertility. One application for manipulation of flower structure can be in the production of saffron, which is derived from the stigmas of *Crocus sativus*.

Closely Related Genes from Other Species

G2143 protein shared extensive homology in the basic helix loop helix region with a protein encoded by Glycine max cDNA clones (AW832545, BG726819 and BG154493) and a *Lycopersicon esculentum* cDNA clone (BE451174). There was lower homology outside of the region.

G2557: G2557 (SEQ ID NO: 133) is a member of a clade of highly related HLH/MYC proteins that also includes G779 (SEQ ID NO: 113), G1063 (SEQ ID NO: 119), G1499 (SEQ ID NO: 7), and G2143 (SEQ ID NO: 129). All of these genes caused similar pleiotropic phenotypic effects when overexpressed, the most striking of which was the production of ectopic carpelloid tissue. These genes can be considered key regulators of carpel development. The flowers of 35S::G2557 primary transformants displayed patches of stigmatic papillae on the sepals, and often had rather narrow petals and poorly developed stamens. Additionally, carpels were also occasionally held outside of the flower at the end of an elongated pedicel like structure. As a result of such defects, 35S::G2557 plants often showed very poor fertility and formed small wrinkled siliques. In addition to such floral abnormalities, the majority of primary transformants were also small and darker green in coloration than wild type. Approximately one third of the T1 plants were extremely tiny and completely sterile. Three T1 lines (#7,9,12), that had produced some seeds, and

showed a relatively weak phenotype, were chosen for further study. All three of the T2 populations from these lines contained plants that were distinctly small, had abnormal flowers, and were poorly fertile compared to controls. Stigmatic tissue was not noted on the sepals of plants from these three T2 lines. Another line (#4) that had shown a moderately strong phenotype in the T1 was sown for only morphological analysis in the T2 generation. T2-4 plants were small, dark green, and produced abnormal flowers with ectopic stigmatic tissue on the sepals, as had been seen in the parental plant. G2557 expression was detected at low to moderate levels in all tissues tested except shoots. It was induced by cold, heat, and salt, and repressed by pathogen infection. G1063 can be used to manipulate flower form and structure or plant fertility. One application for manipulation of flower structure can be in the production of saffron, which is derived from the stigmas of *Crocus sativus*.

Closely Related Genes from Other Species

G2557 protein shows extensive sequence similarity in the region of basic helix loop helix with a protein encoded by Glycine max cDNA clone (BE347811).

G2430: The complete sequence of G2430 (SEQ ID NO: 697) was determined. G2430 is a member of the response regulator class of GARP proteins (ARR genes), although one of the two conserved aspartate residues characteristic of response regulators is not present. The second aspartate, the putative phosphorylated site, is retained so G2430 can have response regulator function. G2430 is specifically expressed in embryo and silique tissue. In morphological analyses, plants overexpressing G2430 showed more rapid growth than control plants at early stages, and in two of three lines examined produced large, flat leaves. Early flowering was observed for some lines, but this effect was inconsistent between plantings. G2430 can regulate plant growth. Overexpression of G2430 in *Arabidopsis* also resulted in seedlings that are slightly more tolerant to heat in a germination assay. Seedlings from G2430 overexpressing transgenic plants were slightly greener than the control seedlings under high temperature conditions. In a repeat experiment on individual lines, G2430 line 15 showed the strongest heat tolerant phenotype. G2430 can be useful to promote faster development and reproduction in plants.

Closely Related Genes from Other Species

G2430 had some similarity within of the conserved GARP and response-regulator domains to non-Arabidopsis proteins.

G1478: The sequence of G1478 (SEQ ID NO: 831) was determined and G1478 was analyzed using transgenic plants in which G1478 was expressed under the control of the 35S promoter. Plants overexpressing G1478 had a general delay in progression through the life cycle, in particular a delay in flowering time. G1478 is expressed at higher levels in flowers, rosettes and embryos but otherwise expression is constitutive. Based on the phenotypes produced through G1478 overexpression, G1478 can be used to manipulate the rate at which plants grow, and flowering time.

Closely Related Genes from Other Species

G1478 shows some homology to non-Arabidopsis proteins within the conserved domain.

G681: G681 (SEQ ID NO: 579) was analyzed using transgenic plants in which the gene was expressed under the control of the 35S promoter. Approximately half of the 35S::G681 primary transformants were markedly small and formed narrow leaves compared to controls. These plants often produced thin inflorescence stems, had rather poorly formed flowers with low pollen production, and set few seeds. Three T1 lines with relatively weak phenotypes, which had produced reasonable quantities of seed, were selected for further study. Plants from one of the T2 populations were noted to be slightly small, but otherwise the T2 lines displayed no consistent differences in morphology from controls. In leaves of two of the T2 lines, overexpression of G681 resulted in an increase in the percentage of the glucosinolate M39480. According to RT-PCR analysis, G681 expression was detected at very low levels in flower and rosette leaf tissues. G681 was induced by drought stress. G681 can be used to alter glucosinolate composition in plants. Increases or decreases in specific glucosinolates or total glucosinolate content are desirable depending upon the particular application. For example: (1) Glucosinolates are undesirable components of the oilseeds used in animal feed, since they produce toxic effects. Low-glucosinolate varieties of canola have been developed to combat this problem. (2) Some glucosinolates have anti-cancer activity; thus, increasing the levels or

composition of these compounds might be of interest from a nutraceutical standpoint.

(3) Glucosinolates form part of a plants natural defense against insects. Modification of glucosinolate composition or quantity could therefore afford increased protection from predators. Furthermore, in edible crops, tissue specific promoters can be used to ensure that these compounds accumulate specifically in tissues, such as the epidermis, which are not taken for consumption.

Closely Related Genes from Other Species

G681 shows some sequence similarity with known genes from other plant species within the conserved Myb domain.

G878: G878 (SEQ ID NO: 611) was studied using transgenic plants in which the gene was expressed under the control of the 35S promoter. Analysis of primary transformants revealed that overexpression of G878 delays the onset of flowering in *Arabidopsis*. 11/20 of the 35S::G878 T1 plants flowered approximately one week later than wild type under continuous light conditions. These plants were also darker green, had shorter stems, and senesced later than controls. G878 was ubiquitously expressed. G878 can be used to modify flowering time and senescence, and a wide variety of applications exist for systems that either lengthen or shorten the time to flowering.

Closely Related Genes from Other Species

G878 was highly related to other WRKY proteins from a variety of plant species, such as the *Nicotiana tabacum* DNA-binding protein 2 (WRKY2) (AF096299), and a *Cucumis sativus* SPF1-like DNA-binding protein (L44134).

G374: G374 (SEQ ID NO: 47) was expressed at low levels throughout the plant and was induced by salicylic acid. G374 was investigated using lines carrying a T-DNA insertion in this gene. The T-DNA insertion was approximately three quarters of the way into the protein coding sequence and should result in a null mutation. Homozygosity for a T-DNA insertion within G374 caused lethality at early stages of embryo development. In an initial screen for G374 knockouts, heterozygous plants were identified. Seed from those individuals was sown to soil and eleven plants were PCR-screened to identify homozygotes. No homozygotes were obtained;

6 of the progeny were heterozygous whilst the other 5 were wild type. This raised the prospect that homozygosity for the G374 insertion was lethal. To examine this possibility further, heterozygous KO.G374 plants were re-grown. These individuals looked wild type, but their siliques were examined for seed abnormalities. When green siliques were dissected, around 25% of developing seeds were white or aborted. Embryos from these siliques were cleared using Hoyers solution, and examined under the microscope. It was apparent that embryos from the white seeds had arrested at early (globular or heart) stages of development, whilst embryos from the normal seeds were fully developed. Such arrested or aborted seeds most likely represented homozygotes for the G374 insertion. To support this conclusion, seed was collected from heterozygous plants and sown to kanamycin plates (the T-DNA insertion carried the NPT marker gene). Of the seedlings that germinated, 160 were kanamycin resistant and 107 were kanamycin sensitive. These data more closely fitted a 2:1 (chi-sq., 1df, = 5.5, $0.05 > P > 0.01$) than a 3:1 (chi-sq., 1df, = 32, $P < 0.001$) ratio. Such a segregation ratio suggested that a homozygous class of kanamycin resistant seedlings was absent from the progeny of KO.G374 plant. G374 can be a herbicide target.

Closely Related Genes from Other Species

Similar sequences to G374 are present in tomato and *Medicago truncatula*, and these sequences can be orthologs.

Example VIII: Identification of Homologous Sequences

Homologous sequences from *Arabidopsis* and plant species other than *Arabidopsis* were identified using database sequence search tools, such as the Basic Local Alignment Search Tool (BLAST) (Altschul et al. (1990) *J. Mol. Biol.* 215:403-410; and Altschul et al. (1997) *Nucl. Acid Res.* 25: 3389-3402). The tblastx sequence analysis programs were employed using the BLOSUM-62 scoring matrix (Henikoff, S. and Henikoff, J. G. (1992) *Proc. Natl. Acad. Sci. USA* 89: 10915-10919).

Identified non-*Arabidopsis* sequences homologous to the *Arabidopsis* sequences are provided in Table 5. The percent sequence identity among these sequences can be as low as 47%, or even lower sequence identity. The entire NCBI GenBank database was filtered for sequences from all plants except *Arabidopsis thaliana* by selecting all entries in the NCBI GenBank database associated with NCBI

taxonomic ID 33090 (Viridiplantae; all plants) and excluding entries associated with taxonomic ID 3701 (*Arabidopsis thaliana*). These sequences are compared to sequences representing genes of SEQ IDs NOs:2 - 2N, where N = 2-561, using the Washington University TBLASTX algorithm (version 2.0a19MP) at the default settings using gapped alignments with the filter "off". For each gene of SEQ IDs NOs:2 - 2N, where N = 2-561, individual comparisons were ordered by probability score (P-value), where the score reflects the probability that a particular alignment occurred by chance. For example, a score of $3.6e-40$ is 3.6×10^{-40} . In addition to P-values, comparisons were also scored by percentage identity. Percentage identity reflects the degree to which two segments of DNA or protein are identical over a particular length. Examples of sequences so identified are presented in Table 5. Homologous or orthologous sequences are readily identified and available in GenBank by Accession number (Table 5; Test sequence ID). The identified homologous polynucleotide and polypeptide sequences and homologues of the *Arabidopsis* polynucleotides and polypeptides may be orthologs of the *Arabidopsis* polynucleotides and polypeptides (TBD: to be determined).

Example IX Introduction of polynucleotides into dicotyledonous plants

SEQ ID NOs:1-(2N - 1), wherein N = 2-561, paralogous, orthologous, and homologous sequences recombined into pMEN20 or pMEN65 expression vectors are transformed into a plant for the purpose of modifying plant traits. The cloning vector may be introduced into a variety of cereal plants by means well-known in the art such as, for example, direct DNA transfer or *Agrobacterium tumefaciens*-mediated transformation. It is now routine to produce transgenic plants using most dicot plants (see Weissbach and Weissbach, (1989) *supra*; Gelvin et al., (1990) *supra*; Herrera-Estrella et al. (1983) *supra*; Bevan (1984) *supra*; and Klee (1985) *supra*). Methods for analysis of traits are routine in the art and examples are disclosed above.

Example X Transformation of Cereal Plants with an Expression Vector

Cereal plants such as corn, wheat, rice, sorghum or barley, may also be transformed with the present polynucleotide sequences in pMEN20 or pMEN65 expression vectors for the purpose of modifying plant traits. For example, pMEN020 may be modified to replace the NptII coding region with the BAR gene of *Streptomyces hygroscopicus* that confers resistance to phosphinothricin. The KpnI

and BglII sites of the Bar gene are removed by site-directed mutagenesis with silent codon changes.

The cloning vector may be introduced into a variety of cereal plants by means well-known in the art such as, for example, direct DNA transfer or *Agrobacterium tumefaciens*-mediated transformation. It is now routine to produce transgenic plants of most cereal crops (Vasil, I., Plant Molec. Biol. 25: 925-937 (1994)) such as corn, wheat, rice, sorghum (Cassas, A. et al., Proc. Natl. Acad Sci USA 90: 11212-11216 (1993) and barley (Wan, Y. and Lemeaux, P. Plant Physiol. 104:37-48 (1994). DNA transfer methods such as the microprojectile can be used for corn (Fromm. et al. Bio/Technology 8: 833-839 (1990); Gordon-Kamm et al. Plant Cell 2: 603-618 (1990); Ishida, Y., Nature Biotechnology 14:745-750 (1990)), wheat (Vasil, et al. Bio/Technology 10:667-674 (1992) ; Vasil et al., Bio/Technology 11:1553-1558 (1993); Weeks et al., Plant Physiol. 102:1077-1084 (1993)), rice (Christou Bio/Technology 9:957-962 (1991); Hiei et al. Plant J. 6:271-282 (1994); Aldemita and Hodges, Planta 199:612-617; Hiei et al., Plant Mol Biol. 35:205-18 (1997)). For most cereal plants, embryogenic cells derived from immature scutellum tissues are the preferred cellular targets for transformation (Hiei et al., Plant Mol Biol. 35:205-18 (1997); Vasil, Plant Molec. Biol. 25: 925-937 (1994)).

Vectors according to the present invention may be transformed into corn embryogenic cells derived from immature scutellar tissue by using microprojectile bombardment, with the A188XB73 genotype as the preferred genotype (Fromm, et al., Bio/Technology 8: 833-839 (1990); Gordon-Kamm et al., Plant Cell 2: 603-618 (1990)). After microprojectile bombardment the tissues are selected on phosphinothricin to identify the transgenic embryogenic cells (Gordon-Kamm et al., Plant Cell 2: 603-618 (1990)). Transgenic plants are regenerated by standard corn regeneration techniques (Fromm, et al., Bio/Technology 8: 833-839 (1990); Gordon-Kamm et al., Plant Cell 2: 603-618 (1990)).

The plasmids prepared as described above can also be used to produce transgenic wheat and rice plants (Christou, Bio/Technology 9:957-962 (1991); Hiei et al., Plant J. 6:271-282 (1994); Aldemita and Hodges, Planta 199:612-617 (1996); Hiei et al., Plant Mol Biol. 35:205-18 (1997)) that coordinately express genes of

interest by following standard transformation protocols known to those skilled in the art for rice and wheat Vasil, et al. Bio/Technology 10:667-674 (1992) ; Vasil et al., Bio/Technology 11:1553-1558 (1993); Weeks et al., Plant Physiol. 102:1077-1084 (1993)), where the bar gene is used as the selectable marker.

All references, publications, patent documents, web pages, and other documents cited or mentioned herein are hereby incorporated by reference in their entirety for all purposes. Although the invention has been described with reference to specific embodiments and examples, it should be understood that one of ordinary skill can make various modifications without departing from the spirit of the invention. The scope of the invention is not limited to the specific embodiments and examples provided.

We claim:

1. A transgenic plant comprising a recombinant polynucleotide having a nucleotide sequence selected from the group consisting of:
 - (a) a nucleotide sequence encoding a polypeptide comprising a sequence selected from those of SEQ ID NOs: 860, 802, 240, 274, 558, 24, 1120, 44, 460, 286, 120, 130, 134, 698, 832, 580, 612, and 48, or a complementary nucleotide sequence thereof;
 - (b) a nucleotide sequence of SEQ ID NOs: 859, 801, 239, 273, 557, 23, 1119, 43, 459, 285, 119, 129, 133, 697, 831, 579, 611, 47, or a complementary nucleotide sequence thereof; and
 - (c) a nucleotide sequence that hybridizes under stringent conditions to a nucleotide sequence of one or more polynucleotides of: (a) or (b).
2. The transgenic plant of claim 1 wherein the transgenic plant possesses an altered trait as compared to another plant, or the transgenic plant exhibits an altered phenotype as compared to another plant, or the transgenic plant expresses an altered level of one or more genes associated with a plant trait as compared to another plant, wherein the other plant does not comprise the recombinant polynucleotide.
3. The transgenic plant of claim 1 wherein the plant possesses an altered trait as compared to another plant wherein the trait is an alteration in one or more physical characteristics selected from the group consisting of: the number of trichomes, fruit and seed size and number, yield of stems, leaves, inflorescences, or roots, stability of seeds during storage, susceptibility of the seed to shattering, root hair length and quantity, internode distances, or the quality of seed coat.
4. The transgenic plant of claim 1 wherein the plant possesses an altered trait as compared to another plant wherein the trait is an alteration in a plant growth characteristic selected from the group consisting of: growth rate, germination rate of seeds, vigor of plants and seedlings, leaf and flower senescence, male sterility, apomixis, flowering time, flower abscission, rate of nitrogen uptake, osmotic sensitivity to soluble sugar concentrations, biomass or transpiration characteristics,

apical dominance, branching patterns, number of organs, organ identity, and organ shape or size.

5. The transgenic plant of claim 1 wherein the plant possesses an altered trait as compared to another plant wherein the trait is an alteration in one or more characteristics selected from the group consisting of protein or oil production, seed protein or oil production, insoluble sugar level, soluble sugar level, and starch composition.
6. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:860.
7. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:802.
8. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:240.
9. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:274.
10. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:558.
11. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:24.
12. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:1120.
13. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:44.

14. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:460.
15. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:286.
16. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:120.
17. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:130.
18. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:134.
19. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:698.
20. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:832.
21. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:580.
-
22. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:612.
23. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:48.
24. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:859.

25. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:801.
26. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:239.
27. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:273.
28. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:557.
29. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:23.
30. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:1119.
31. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:43.
32. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:459.
33. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:285.
34. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:119.
35. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:129.

36. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:133.
37. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:697.
38. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:831.
39. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:579.
40. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:611.
41. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:47.
42. The transgenic plant of claim 1, further comprising a constitutive, inducible, or tissue-specific promoter operably linked to said nucleotide sequence.
43. The transgenic plant of claim 1, wherein the plant is selected from the group consisting of: soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower, alfalfa, sugarcane, turf, banana, blackberry, blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, pumpkin, spinach, squash, sweet corn, tobacco, tomato, watermelon, mint and other labiates, rosaceous fruits, and vegetable brassicas.
44. The transgenic plant of claim 1 wherein the encoded polypeptide is expressed and regulates transcription of a gene.

45. A method of using the transgenic plant of claim 1 to grow a progeny plant from a parent plant, the method comprising crossing the transgenic plant with another plant, selecting seed, and growing the progeny plant from the seed.
46. An isolated or recombinant polynucleotide comprising a nucleotide sequence selected from the group consisting of:
- (a) a nucleotide sequence encoding a polypeptide comprising a sequence selected from SEQ ID NOs: 240, 274, 558, 286, 698, and 832, or a complementary nucleotide sequence thereof;
 - (b) a nucleotide sequence of SEQ ID NOs: 239, 273, 557, 285, 697, 831, or a complementary nucleotide sequence thereof; and
 - (c) a nucleotide sequence that hybridizes under stringent conditions to a nucleotide sequence of one or more of: (a) or (b).
47. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO: 240.
48. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO: 274.
49. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO: 558.
50. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO: 286.
51. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO: 698.

52. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:832.
53. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:239.
54. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:273.
55. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:557.
56. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:285.
57. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:697.
58. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:831.
59. The isolated or recombinant polynucleotide of claim 46, further comprising a constitutive, inducible, or tissue-specific promoter operably linked to the nucleotide sequence.
60. The isolated or recombinant polynucleotide of claim 46 wherein the encoded polypeptide is expressed and regulates transcription of a gene.
61. A vector comprising the isolated or recombinant polynucleotide of claim 46.
62. A host cell comprising the vector of claim 61.

63. A method of using the isolated or recombinant polynucleotide of claim 46 for producing a plant having a modified trait, the method comprising selecting a polynucleotide that encodes a polypeptide, inserting the polynucleotide into an expression vector, introducing the vector into a plant or a cell of a plant to overexpress the polypeptide, thereby producing a modified plant, and selecting a modified plant for a modified trait.
64. The method of claim 63 wherein the plant possesses a modified trait as compared to another plant wherein the trait is an alteration in one or more physical characteristics selected from the group consisting of: the number of trichomes, fruit and seed size and number, yield of stems, leaves, inflorescences, or roots, stability of seeds during storage, susceptibility of the seed to shattering, root hair length and quantity, internode distances, or the quality of seed coat.
65. The method of claim 63 wherein the plant possesses a modified as compared to another plant wherein the trait is an alteration in a plant growth characteristic selected from the group consisting of: growth rate, germination rate of seeds, vigor of plants and seedlings, leaf and flower senescence, male sterility, apomixis, flowering time, flower abscission, rate of nitrogen uptake, osmotic sensitivity to soluble sugar concentrations, biomass or transpiration characteristics, apical dominance, branching patterns, number of organs, organ identity, and organ shape or size.
66. The method of claim 63 wherein the plant possesses a modified trait as compared to another plant wherein the trait is an alteration in one or more characteristics selected from the group consisting of protein or oil production, seed protein or oil production, insoluble sugar level, soluble sugar level, and starch composition.
67. A modified plant produced by the method of claim 63.
68. A method of using the plant of claim 67 to grow a progeny plant from a parent plant, the method comprising crossing the transgenic plant with another plant, selecting seed, and growing the progeny plant from the seed.

69. The plant produced by the method of claim 68.

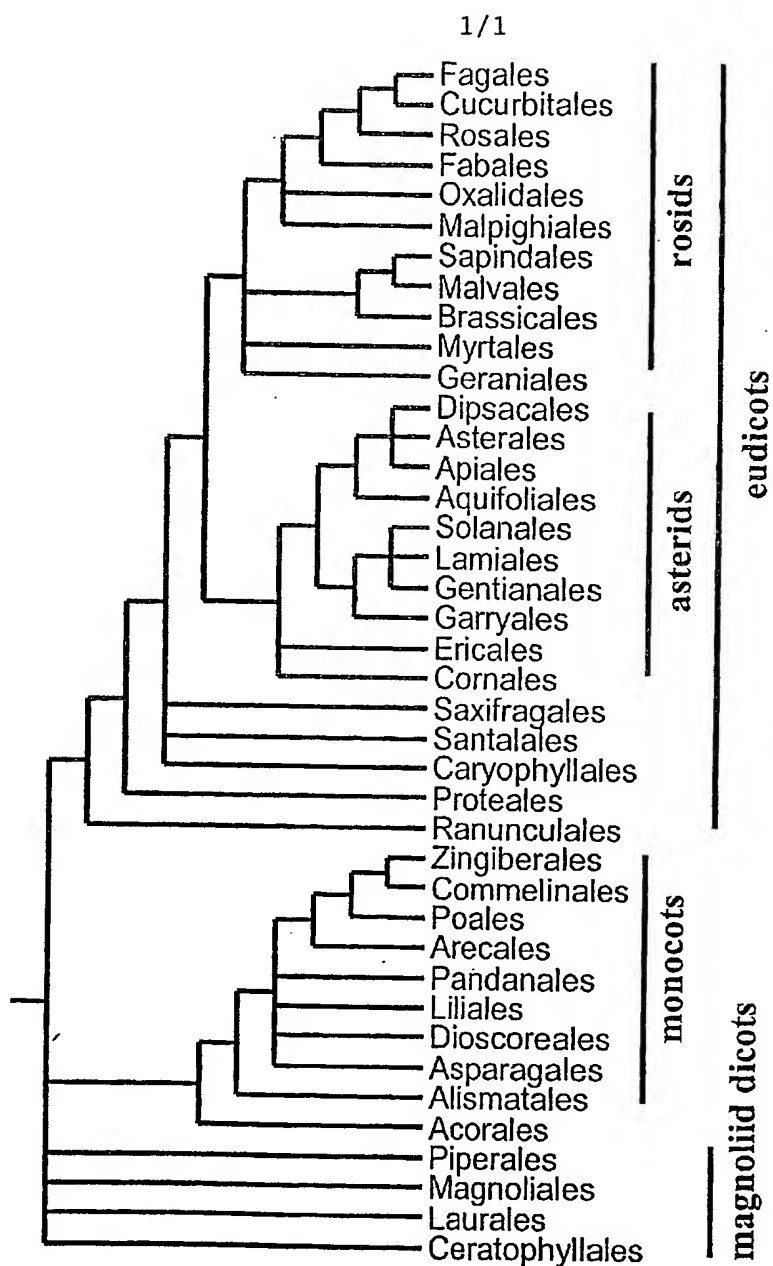


Figure 1

SEQUENCE LISTING

<110> Mendel Biotechnology, Inc.

Ratcliffe, Oliver
 Riechmann, Jose Luis
 Adam, Luc J.
 Dubell, Arnold T.
 Heard, Jacqueline E.
 Pilgrim, Marsha L.
 Jiang, Cai-Zhong
 Reuber, T. Lynne
 Creelman, Robert A.
 Pineda, Omaira
 Yu, Guo-Liang
 Broun, Pierre E.

<120> YIELD-RELATED POLYNUCLEOTIDES AND
POLYPEPTIDES IN PLANTS

<130> 514442002041

<150> 60/310,847

<151> 2001-08-09

<150> 60/336,049

<151> 2001-11-19

<150> 60/338,692

<151> 2001-12-11

<150> 10/171,468

<151> 2002-06-14

>G1275 (58..579)

CCAAGAAAAGGGAAGATCACGCATTCTTATAGGCGTAATTCGTAAATAGTGGTGAGTATG
 AATGATGCAGACACAACTTGGGGAGTAGTTTCAGCGATGATACTCACTCTGTGTTTCGAG
 TTTCCGAGCTAGACTTGTCTAGATGAATGGATGGATGATGATCTTGTGTCTGCGGTTTCC
 GGGATGAATCAGTCTTATGGTTATCAGACTAGTGATGTTGCTGGTGCTTTATTCTCAGGT
 TCTTCTAGCTGTTTCAGTCATCCTGAATCTCCAAGTACCAAACTTATGTTGCTGCTACA
 GCCACTGTTCTGCCGACAACCAAAACAAGAAAAGAAAAGAAAAAATTAAAGGGAGAGTT
 GCGTTCAAGACACGGTCCGAGGTGGAAGTGCTTGACGACGGGTTCAAGTGGAGAAAGTAT
 GGAAGAAGATGGTGAAGAACAGCCACATCCAAGAACTACTACAAATGTTTCAGTTGAT
 GGCTGTCCCGTGAAGAAAAGGGTTGAACGAGACAGAGATGATCCGAGCTTTGTGATAACA
 ACTTACGAGGGTTCCCAATCACTCAAGCATGAACCTAAGACTCGAACTAAGGCTCAAGG
 CGACCATGCTATATTCAGCACATCTTATTTCTATGGTTACGAACGATACTTAAACTGC
 TTCTAGTTCTTTATATCCATTGTAACTGGTTGCAGGTTACAAAATTTTGAGAGGTTTAT
 GACATTCTAAATCTGTAGTACTTATATA

>G1275 Amino Acid Sequence (domain in AA coordinates: 113-169)

MNDADTNLGSSFSDDTHSVFEFPELDLSDEWMDDDLVSAVSGMNQSYGYQTSVAGALFS
 GSSSCFSPHSPSTKTYVAATATASADNQNKKKKIKGRVAFKTRSEVEVLDDGFKWRK
 YGKKMVKNSPHPRNYKCSVDGCPVKRVERDRDDPSFVITTYEGSHNHSSMN*

>G1411 (110..856)

TAAAGAAAACTGAACAACCCTAAAGTACTGTATAAATCCTATATCAAATTTTTTTTGTG
 GAAGAAAAGGCTATATTTAAAAGAAAATCAAGCAAAAGTAGATCCTCGGATGTATGGGAA
 GAGGCCCTTTGGAGGTGATGAATCTGAAGAAAGGGAAGAAGATGAGAAGTTGTTCCCGGT
 CTTCTCGGCCCGATCTCAACACGACATGCGTGTTATGGTCTCGGCCCTGACTCAAGTAAT

CGGAAACCAACAAAGCAAATCTCATGATAACATCAGCTCTATTGATGATAACTATCCTTC
 TGTGTATAATCCACAAGACCCTAATCAACAAGTTGCGCCTACTCATCAAGACCAAGGGGA
 CTTGAGGAGGAGACATTATAGAGGTGTAAGGCAAAGGCCATGGGGAAGTGGGCAGCTGA
 AATCCGAGACCCAAAAAAGGCGGCACGTGTGTGGCTCGGGACATTGAAACCGCTGAATC
 TGGCGCCTTAGCTTATGATGAAGCAGCCCTAAAGTTCAAAGGAAGCAAAGCAAACTCAA
 TTTCCCGGAGAGGGTTTCAGCTTGGAGTAACCTACATATTACTCCTCCAACCAAAATCC
 ACAAATGGAACCACAAAGTATACCGAACTATAATCAATACTATCATGATGCGAGTAGTGG
 TGATATGCTAAGTTTTAATTTGGGCGGTGGGTATGGGAGTGGTACCGGATATTCAATGTC
 TCATGATAATAGTACTACGACTGCTGTACAACCTTCTTCGTCTTCTGGTGGCTCTTCTAG
 GCAACAAGAAGAGCAAGATTATGCCAGATTCTGGCGCTTTGGGGATTCTTCTTCTCTCC
 TCATTCGGGATATTAAATTAGGAGATTGATCAGTTACTTGTGATGAAGTAATGATACATT
 TCCCGTCAAAATTGAGATGATCATATGCTTCTGAAATGTTTTTGTAGTGTCATTTTTGTCT
 TCCCGCTTAAGATTTATGAACGTGTTTTCTTGTTTTTTGGTTAAAAA
 AAAAAAAAAAAAAA

>G1411 Amino Acid Sequence (domain in AA coordinates: 87-154)
 MYGKRPFQDESEEREEDENLFPVFSARSQHDMRVMVSALTQVIGNQQSKSHDNISSIDD
 NYPSVYNPDQPNQOVAPTHQDQGLRRRHRYRGVRQRPWGKWAAEIRDPKKAARVWLGTFE
 TAESAALAYDEAALKFKGSKAKLNFPERVQLGSNSTYYSSNQIPQMEPQSI PNYNQYYHD
 ASSGDMLSFNLGGYGS GTGYSM SHDNSTTTAATSSSSGGSSRQEEQDYARFWRFGDS
 SSSPHSGY*

>G1488 (1..996)
 ATGGAAGATGAAGCACATGAATCTTCCACACATCTGATTTTGCCGTTGATGACCTTTTA
 GTTGATTTCTTAACGATGATGACGAAGAAAACGATGTTGTTGCTGATTCACCACTACC
 ACCACCATAACCGACAGCTCTAACTTCTCCGCTGCTGATCTTCCAGTTTCCACGGTGAT
 GTTCAAGACGGCACTAGCTTCTCCGGTGACCTTTGTATACCTTCTGATGATTTGGCTGAT
 GAGTTAGAGTGGCTTTTCGAACATTTGTGGATGAATCATTGTGCGCTGAAGATGTACACAAG
 CTCGAGCTAATATCCGGTTTTAAGAGTCGACCGGACCCGAAATCCGATACCGGAAGCCCG
 GAAAACCCGAATAGCAGCAGTCCGATTTTTACTACCGACGTTTCTGTACCGGCCAAAGCT
 AGAAGCAAACGCTCACGCGCCGCTGCGTGTAATTGGGCCTCACGTGGGCTTCTCAAGGAA
 ACGTTTTACGACAGTCCCTTTCACCGGAGAAACATTCTCTCTAGCCAACAACACTTGTCT
 CCGCCAACCTCGCCGCTTTGTTGATGGCTCCGCTAGGGAAAAAGCAAGCCGTTGATGGA
 GGACACCGACGGAAGAAGGATGTTCTTCACCGAGTCTGGTGGCGCAGAGGAGAGACGG
 TGTCTCCACTGCGCCACGGATAAGACTCCGCAATGGCGGACAGGCCCAATGGGCCGAAG
 ACGTTGTGCAACGCTTGCGGTGTTAGGTACAAATCGGGACGTTTAGTGCCGGAGTATCGG
 CCCGCGGCGAGTCCGACGTTTGTGCTGGCGAAACACTCAAATCTCATCGGAAAGTTATG
 GAGCTCCGCGACAGAAGGAGATGAGTAGGGCCCATCATGAGTTCATACATCACCATCAC
 GGTACGGACACTGCCATGATTTTCGACGTTTCATCGGACGGTGATGATTACTTGATCCAC
 CACAACGTTGGCCAGATTTTCAGACAGCTTATTTGA

>G1488 Amino Acid Sequence (domain in AA coordinates: 221-246)
 MEDEAHEFFHTSDFAVDDLLVDFSNDDDEENDVVADSTTTTTITDSSNFSAADLPSFHGD
 VQDGTFSFGDLCLIPSDDLADELEWLSNIVDESLSPELVHKLLEISGFKSRPDPKSDTGSP
 ENPNSSSPIFTTDSVPAKARSKRSRAAACNWSRGLLKETFYDSPFTGETILSSQQHLS
 PPTSPPLLMAPLGKKQAVDGGHRRKKDVSSPESGAEERRCLHCATDKTPQWRTGPMGPK
 TLCNACGVRYKSGRLVPEYRPAASPTFVLAKHSNSHRKVMELRRQKEMSRAHHEFIHHHH
 GTDTAMIFDVSSDGDYLIHNVGPDFRQLI*

>G1499 (159..833)
 TCGACTCCTTAATTGCATCACCAACCTAACCCCTCATCATATGATCACCTTTCTTTTTT
 CCTTTAATATATATATATGCTCACACACACATATATATATACATATAAGCATCGCCTC
 AAGCATTAAAATTTTACGAACCAAAACAAAAATATGAATAATTATAATATGAACC
 CATCTCTCTTCCAAAATTACACTTGGAAACATCATCAACAGCAGCAACAACAACA
 AGAATGATGATCATCATCATCAACATAATAATGATCCAATCGGTATGGCCATGGACCA
 ACACACAGCTCCATATCTTCAATCCTTTCTCTTCTCATTTCCTCTCTCTCTCTCT
 CCCTCACAAACCACACTTCTCTCCGGAGATCAAGAAGACGACGAAGACGAAGAAGAAC
 CTCTAGAGGAACCTCGGTGCTATGAAGGAATGATGTACAAGATCGCAGCCATGCAATCGG
 TTGACATCGACCCAGCAACCGTCAAGAAACCCAAACGCCGTAACGTGAGGATCTCCGACG
 ACCCTCAGAGTGTGGCGGCTAGACATCGCCGTGAGAGAATCAGTGAGAGGATCAGAAATC
 TTCAGAGACTCGTGCCAGGTGGCACTAAATGGATACGGCTTCAATGCTCGATGAAGCTA

TACGCTATGTCAAGTTCTTGAAACGGCAGATCCGGCTACTCAATAATAATACCGGATATA
CTCTCCGCCGCCGCAAGATCAAGCTTCTCAGGCGGTGACGACGTCATGGGTTTCACCGC
CACCACCGCCAAGTTTCGGCCGTGGGGGAAGAGGAGTAGGAGAATTAATCTAGACAAGAT
GACATTTCCATTAGTAGTAATAATTATGCTATAATGTGTGAGTAATGGTGAATTATG
GA

>G1499 Amino Acid Sequence (domain in AA coordinates: 118-181)
MNNYNMNP SLFQNYTWNNIINSSNNNNKNDDHHHQHNNDPIGMAMDQYTLHIFNPFSSS
HFPPLSSSLTTLTLLSGDQEDDEDEEEPLEELGAMKEMMYKIAAMQSVDIDPATVKKPKR
RNVRI SDDPQSVAAHRHRERISERIRILQRLVPGGTKMDTASMLDEAIRYVKFLKRQIRL
LNNNTGYTPPPPDQASQAVTTSWVSPPPPPSFGRGGRGVGELI*

>G1543 (1..828)

ATGATAAACTACTATTTACGTACATATGCACATACACATATAAACTATATGCTCTATAT
CATATGGATTACGCATGCGTGTGTATGTATAAATAAAGGCATCGTCACGCTTCAAGTT
TGTCCTCTTTTATATTAACCTGAGAGTTTTCTCTCAAACCTTACCTTTTCTTCTCGATC
CTAGCTCTTAAGAACCCTAATAATTCATTGATCAAAATAATGGCGATTTTGCCGGAAGAAC
TCTTCAAACCTTGGATCTTACTATCTCCGTTCCAGGCTTCTCTTCATCCCTCTCTCGGAT
GAAGGAAGTGGCGGAGGAAGAGACCAAGCTAAGGCTAGACATGAATCGGTTACCGTCGTCT
GAAGACGGAGACGATGAAGAATTCAGTCACGATGATGGCTCTGCTCCTCCGCGAAGAGAAA
CTCCGCTCAACAGAGAACAGTCACGCTCTTCTGAAGATAGTTTCAGACAGAATCATACC
CTTAATCCCAAACAAAGGAAGTACTTGCCAAGCATTGATGCTACGGCCAAGACAAATT
GAAGTTTGGTTTCAAACCGTAGAGCAAGGAGCAAATTGAAGCAAACCGAGATGGAATGC
GAGTATCTCAAAGGTGGTTTGGTTTCAATACGGAAGAAAACCAAGGCTCCATAGAGAA
GTAGAAGAGCTTAGAGCCATAAAGGTTGGCCCAACAACGGTGAACCTGCGCTCGAGCCTT
ACTATGTGTCTCTGCTGCGAGCGAGTTACCCCTGCCGCGAGCCCTTCGAGGGCGGTGGTG
CCGGTCCCGCTAAGAAAACGTTTCCGCCGCAAGAGCGTGATCGTTGA

>G1543 Amino Acid Sequence (domain in AA coordinates: 135-195)
MIKLFTYICTYTYKLYALYHMDYACVCMYKYKGIIVTLQVCLFYIKLRVFLSNFTFSSSI
LALKNPNNSLIKIMAILPENSSNLDLTISVPGFSSSPLSDESGSGGRDQLRLDMNRLPSS
EDGDDEEFSSHDDGSAPPRKKRLRLTREQSRLLDSFRQNHTLNPQKEVLAKHLMRLPRQI
EVWFQNRARRARSLKQTEMECEYLKRWFGSLTEENHRLHREVEELRAIKVGPTTVNSASSL
TMCPRCERVTPAASPSRAVVPVPAKKTFFPPQERDR*

>G1635 (1..1164)

ATGGCGTCTGCTCCGTTGACTGCAAATGTTTCAGGGTACCAACGCTTCTTTGAGGAATAGA
GATGAAGAACTGCAGACAAGCAGATACAATTCAATGACCAAAGTTTTGGGGGAAATGAC
TATGCACCCAAGGTACGGAAGCCATACACGATAACAAAAGAGAGAGAGATGGACAGAT
GAAGAGCACAAGAAGTTTGTGAAGCCTTGAAATTATACGGGCGAGCTTGGAGACGAATA
GAAGAACATGTGGGCTCAAAGACCGCAGTTCAGATTGCAAGCCATGCTCAGAAGTTTTTC
TCTAAGGTTGCTCGAGAAGCAACTGGAGGTGATGGGAGCTCAGTAGAGCCGATTGTAATA
CCTCTCTCTCGTCCCAAGAGAAAGCCAGCGCATCCGTACCTCGTAAGTTTGGGAACGAG
GCAGATCAAACAAGTAGATCGGTTTCTCCCTCAGAACGTGATACTCAATCTCCAACCTCT
GTGTTGTCTCACTGTTGGATCAGAAGCATTGTGTTCCCTTGATTGAGTTACCCCAATCGA
AGCTTGTCCCAAGTTTCTTCTGCATCACCACAGCTGCTCTTACAACCACTGCAAATGCA
CCTGAAGAGCTTGAGACTCTGAAGCTGGAGTTGTTTCTTAGTGAGAGACTCTTAAACAGG
GAGAGCTCGATCAAGGAACCAACGAAGCAAAGTCTTAAACTCTTTGGGAAGACAGTTTTG
GTATCTGATTCAAGCATGTCTCTTCTCTAACAACCTCAACATATTGTAAATCCCAATT
CAGCCATTACCAGGAACTCTCATCATCCAAGACACTACCCATAATAAGAACTCACAA
GAAGAAGCTCTGAGCTGCTGGATACAAGTCCCTCTTAAGCAAGAAGATGTGGAAAATAGA
TGTTTGGATTGAGAAAGGCTGTCCAAAACGAAGGATCATCGACTGGATCAAACACTGGT
TCGGTGGATGATACGGGACACACGGAAAGACCACAGAACCCGAAACAATGCTATGTCAA
TGGGAGTTTAAACCAAGTGAGAGGCTGTCATTTTCTGAGCTCAGAAGAACAACCTCCGAG
TCAAATTCAGAGGATTTGGTCCATACAAGAAGAGAAAGATGGTAACAGAAGAAGAAGAG
CATGAGATTCATCTCCACTTATAA

>G1635 Amino Acid Sequence (domain in AA coordinates: 44-104)
MASSPLTANVQGTNASLRNDEETADKQIQFNDQSFSGNDYAPKVRKPYTITKERERWTD
EEHKKFVEALKLYGRAWRRIEEHVGSKTAVQIRSHAQKFFSKVAREATGGDGSSVEPIVI
PPRPKRKPAHPYPRKFGNEADQTSRVSPPSERDTQSPTSVLSTVGSEALCSLDSSSPNR
SLSPVSSASPPAALTTTANAPEELETLKLELFPSERLNRESSIKEPTKQSLKLFKTVL

VSDSGMSSSLTTSTYCKSPIQPLPRKLSSSKTLPIIRNSQEELLSCWIQVPLKQEDVENR
CLDSGKAVQNEGSSTGSNTGSDDTGHTTEKTTEPETMLCQWEFKPSERSAFSELRRTNSE
SNSRGFGPYKKRKMVTEEEEHEIHLHL*

>G1794 (160..1335)

TCTTTCTTTCTTCTCTTTGTCTCTGTTTCTGTTTCTCTCTCTCTCTCTACAGAGTT
TTCTTTCCCTCGAAGAAAAAGAATATTTTAAATTTAATTTCTCTGCGTTTATAAGCTT
TAAGTTTCAGAGGAGGAGGATTTAGAAGGAGGTTTTGTATGTGTCTTAAAAGTGGCA
AATCAGGAAGATAACGTTGGCAAAAAGCCGAGTCTATTAGAGACGATGATCATCGGACG
TTATCTGAAATCGATCAATGGCTTTACTTATTCGCAGCCGAAGACGACCACCGTCAT
AGCTTCCCTACGCAGCAGCCGCTCCATCGTCGTCGTCCTCATCTCTTATCTCAGGTTTC
AGTAGAGAGATGGAGATGTCTGCTATGTCTCTGCTTTGACTCACGTTGTTGCTGGAAAT
GTTCTTCAGCATCAACAAGGAGGCGGTGAAGGTAGCGGAGAAGGACTTCGAATTCGTCCT
TCTTCTTCGGGGCAGAAAAGGAGGAGAGAGGTGGAGGAAGGTGGCGCCAAAGCGTTAAG
GCAGCTAATACTTTGACGGTTGATCAATATTTCTCCGGTGGTAGCTCTACTTCTAAAGTG
AGAGAAGCTTCGAGTAACATGTCTAGGTCGGGGCCCAACATACGAGTATACAACCTACGGCA
ACTGCTAGTAGCGAAACGTCGTCGTTTGTAGTGGGGACCAACCTCGCGAAGATACAGAGGA
GTTAGACAAAGACCATGGGGAAAGTGGGCGGCTGAGATTCGAGATCCATTTAAAGCAGCT
AGAGTTTGGCTCGGTACGTTTCGACAATGCTGAATCAGCAGCAAGAGCTTACGACGAAGCT
GCACTTCGGTTTATAGAGGCAACAAAGCCAACTCAACTTCCCTGAAAACGTCAAACCTCGTT
AGACCTGCTTCAACCGAAGCACAACCTGTGCACCAACCGCTGCTCAAAGACCGACCCAG
TCAAGGAACCTCGGTTCAACGACTACCTTTTGCCCATAGACCTGCTTCGAATCAAAGC
GTTTATTTCGCAGCCGTTGATGCAATCATACAACCTTGAGTTACTCTGAAATGGCTCGTCAA
CAACAACAGTTTCAGCAACATCATCAACAATCTTTGGATTATACGATCAAATGTCGTTT
CCGTTGCGTTTTCGGTCACTGGAGGTTCAATGATGCAATCTACGTCGTCATCATCTCT
CATTCTCGTCTCTGTTTTCCTCCGGCTGCTGTTTCAGCCGCCACCAGAATCAGCTAGCGAA
ACCGGTTATCTCCAGGATATACAATGGCCATCAGACAAGACTAGTAATACTACAATAAT
AGTCCATCCTCCTGATGACTTGCTTCATTTTATTTGTTTCTACTATAGAGTAATAGAAAAC
AGGAAAATGATTATATGTTATAGAGTTATTTTCCAAATATTATAGGGTTTAGGTTGTTT
GTATTGTTCTGCTTTCATCCTCTCATGCTTTTTTTCTTAATTTATTATTTTGCATTA
TAATTTGTTTTCATTGTAACAAACATTAAAAAGACCACATGGAGAAAGGAAAAAAGAG
AG

>G1794 Amino Acid Sequence (domain in AA coordinates: TBD)

MCVLKVANQEDNVGKKAESIRDDHRTLSEIDQWLYLFAAEDDHRHSFPTQQPPSSSS
SSLISGFSREMESAIVSALTHVAGNVPOHQGGGEGSGEGTSNSSSSSGQKRREVEE
GGAKAVKAANTLTVDQYFSGGSSTSKVREASSNMSGPGPTYEYTTATASSETSSFSGDQ
PRRRYRGVRQRPWGKWAIEIRDPFKAAARVWLGTFDNAESAARAYDEAALRFRGNKAKLNF
PENVKLVRPASTEAPVHQTAQRPTQSRNSGSTTTLLPIRPASNQSVHSQPLMQSYNLS
YSEMARQQQQFQHHQQLDLYDQMSFPLRFHGTGGSMQSTSSSSSHSRPLFSPAAPVQ
PPESASETGYLQDIQWPDKTSNNYNNSPSS*

>G1839 (38..592)

ATCACAGTTATGTTTCCATTTCATTGGCTATAAAAACCATGCTCACTCCCTTTTGTCTTC
ACACCATTTCAGGAAAAAATGAATAGTTGTCTAGTCTAATCCCACCAAAATGGATAATTC
AGAAAATGTTCTATTTAATGATCAAAACGAAAATTTCACTTGTTCACCAACACCCCTTC
TTCTTCGTACTTGACAAGAGATCAAGAGCACGAGATCATGGTCTCTGCTCTGCGACAAGT
GATATCTAATCCCGAGCTGACGACGCTCATCAAACTTGATCATCACAAGCGTTCC
GCCTCCAGACGCTGGCCCTTGTCTCTCTGTGGCGTCGCCGTTGCTACGGCTGCACATT
ACAACGGCCGACCGAGAGGTAAAGAAGGAGAAGAAATACAAAGGAGTAAGGAAAAAACC
ATCGGGTAAATGGGCGGCGGAGATATGGGATCCGAGATCGAAATCAAGGAGGTGGCTTGG
AACGTTTCTTACGCGCGAGATGGCGGCACAATCTTACAATGATCGGGCGGCTGAGTATCG
AGCAAGACGTGGTAAACAAACGGAGAAGGAATTAACGGCGGTGGAGATGACTGAGAAG
GACATGGTGGTGATCATACACGGCGAGGTGGAAATGTTATATTACTATTGAAAACTAA
ATTATTTATTATAGAGGGAGATATTACTCTTTACGCTTTTCATTAAAGATTATTTTATAA
GTTTTAAAGTATTTTATTGTTATAAAAAA

>G1839 Amino Acid Sequence (domain in AA coordinates: TBD)

MLTPFCSSHLQEKMNCSQSNPTKMDNSENVLFNDQENFTLVAPHPSSSYLTRDQEHEI
MVSALRQVISNSGADDASSNLIITSVPPPDAGPCPLCGVAGCYGCTLQRPHEVKEKK
YKGVKKPSGKWAIEIWDPRKSRRWLGTFLTAEMAAQSYNDAAEYRARRGKTNGEGIK

RRWR*

>G2108 (35..694)

GAGAGAGAAACATTGATCTCTGAATATTGTGAACATGTTGAAATCAAGTAACAAGAGAAA
AAGCAAAGAAGAGAAGAAGTTACAAGAAGGGAAGTACCTTGGAGTGAGGAGACGTCCATG
GGGAAGATATGCAGCTGAAATCAGAAACCCCTTTACTAAAGAAAGACATTGGCTTGAAC
GTTTGATACAGCCGAAGAAGCTGCTTTGCATATGACGTTGCTGCTCGATCCATCAGCGG
CTCTCTAGCTACAAACAACTTCTTCTACACTGAAACACCTCTTTAGAAAGACATCCACA
ACAGTCTTTGGAGCCTCATATGACTTGGGGATCTTCTAGTCTCTGCTTCTTCAAGATCA
GCCTTTTGAAAACAACCATTTTGTGCTGATCCTATCTCTTCTTCTTTTCTCAAAAACA
AGAGTCTTCTACCAATCTCACTAACACTTTCTCACATTGTTATAATGATGGTGATCATGT
TGGCCAAAGCAAAGAGATTTCTTTACCTAATGATATGTCAAACAGTTTATTTCGGTCATCA
GGACAAAGTCGGTGAACATGACAATGCAGACCATATGAAGTTTGGCTCAGTTCTCAGCGA
CGAACCTCTCTGCTTTGAGTACTACATTGGGAATTATCTTCAGAGTTTCTCAAAGA
TGTC AACGACGATGCTCCACAGTTTCTTATGTGAGCTTGTATTACCGATCCTTCAATTTA
TG

>G2108 Amino Acid Sequence (domain in AA coordinates: 18-85)
MLKSSNKRKSKEEKLQEGKYLGVRRRPWGRYAAEIRNPFTKERHWLGTFTAEAAAFAY
DVAARSISGSLATTNFFYTENTSLEHPPQSLEPHMTWGSSSLCLLDQDPFENNHFVADP
ISSFSQKQESSTNLNTNFSHCYNDGDHVGQSKEISLPNDMSNSLFGHQDKVGEHDNADH
MKFGSVLSDEPLCFEYDYIGNYLQSF LKDVND DAPQFLM*

>G2291 (27..797)

GCTTTCTCACCTTTATAAAATAGAAAATGGAAAACAGCTACACCGTTGATGGTCACCGTC
TTCAATATTCCGTTCCGTTAAGCTCCATGCATGAAACCAGTCAAAACTCCGAAACTTACG
GATTATCCAAAGAGTCGCCGTTGGTCTGCATGCCCTTGTTCGAAACCAACTACTTCAT
TCGATATCTCTTCTCTTTTCTCGTTTAACCCAAAACAGAACCCGAAACACGCATCGTG
TCATGGACGATTCCATCGCCCGCTCGTGGGCGAAAACGTTCTTTTCGGTGATAAAAACA
AAGTCTCTGATCACTTGACCAAAGAAGGTGGTGTGAAGCGGGGGCGAAGATGCCGCAGA
AGACCGGAGGATTCATGGGAGTGAGAAAACGGCCGTGGGGGAGATGGTCGGCGGAGATAA
GAGACAGGATAGGGCGGTGCAGACACTGGTTAGGAACGTTTCGACACGGCGGAAGAGGCAG
CGCGTGCGTATGACGCGGCGGCGAGGAGGCTTAGAGGGACCAAAGCCAAGACCAATTTTCG
TGATTCTCCGCTTTTCCCAAGGAAATAGCTCAGGCTCAGGAGGATAATAGGATGAGGC
AGAAGCAGAAGAAGAAGAAGAAAAAGTGAGTGTGAGGAAGTGTGTTAAAGTCACAT
CGGTTGCACAGTTGTTCGATGATGCCAATTTTATAAATTCTTCTAGTATTAAAGGAAATG
TGATTAGTTCTATTGATAATCTTGAAAAAATGGGTCTAGAGCTTGATTGAGTTAGGGT
TGTTGTCTAGGAAGTGATAAAGCACTCGTAGTTAAGTAGTTGTAGTT

>G2291 Amino Acid Sequence (domain in AA coordinates: TBD)
MENSYTVDGHRLQYSVPLSSMHETSQNSETYGLSKESPLVCMPLFETNTTSFDDISSLFSS
NPKPEPENTHRVMDSDIAAVVGENVLFPGDNKVS DHLTKEGGVKRGRKMPQKTGGFMGVR
KRPWGRWSAEIRDRIGRCRHWLGTFTAEBAARAYDAAARRLRGTKAKTNFVIPPFPKE
IAQAQEDNMRMQKQKKKKKVSVRKCVKVTSAQLFDDANFINSSSIKGNVISSIDNLE
KMGLELDLSLGLLSRK*

>G2452 (1..804)

ATGTCATCGTCGACGATGTACAGAGGAGTTAATATGTTTTCACCGGCAAACACAACTGG
ATTTTTCAAGAAGTCAGAGAAGCCACGTGGACGGCGGAGGAGAACAACCGTTTCGAGAAA
GCTCTCGCTTATCTGGACGACAAAGACAATCTTGAGAGCTGGTCCAAGATCGCAGATTG
ATTCGCGGCAAAACAGTAGCTGACGTCATTAAACGATACAAGGAGCTAGAGGATGATGTC
AGCGACATCGAAGCCGGACTTATCCCATTCGCGGATACGGCGGCGACGCCTCTCCGCT
GCAAACAGTGACTATTTCTTTGGTCTAGAAAACTCCAGCTACGGTTATGATTACGTCGTT
GGAGGAAAGAGGAGTTCCGCGGCGATGACTGATTGTTTTAGGTCTCCGATGCCGGAAG
GAGAGGAAGAAAGGAGTTCCGTGGACCGAGGACGAACACCTACGATTTCTGATGGGTTTG
AAGAAATATGAAAAGGAGATTGGAGAAACATAGCAAAAAGCTTTGTGACGACTCGAACG
CCGACGCAAGTCGCTTACACGCTCAGAAATATTTTCTTCGACAACCTACAGATGGTAAA
GACAAAAGACGATCAAGTATTCACGATATCACCAGTTTAACATCCCTGACGCAGACGCA
TCCGCAACCGCCACGACCGCTGACGTAGCACTCTCTCCTACTCCAGCCAATTTCTTTGAC
GTTTCTCTTCAGCCAATCTCATTACAGTTTCGCGTCTGCGTCTAGCTATTAT
AATGCGTTTTCCGAGTGGAGTTAA

>G2452 Amino Acid Sequence (conserved domain in AA coordinates: 27-213)

MSSSTMYRGVNMFS PANTNWI FQEVREATWTA EENKRF EKALAYLDDKDNLESWSKIADL
IPGKTVADVIKRYKELEDDVSDIEAGLIPIPGYGGDASSAANS DYFFGLENSSYGYDYV
GGKRSSPAMTDCFRSPMPEKERKKGVPTWTEDEHLRFLMGLKKYKGKGDWRNIAKSFVTTTRT
PTQVASHAQKYFLRQLTDGKDKRRSSIHDITTVNI PDADASATATTADVALSPTPANSFD
VFLQPNPHYSFASASASSYNAFPQWS*

>G2509 (143..934)

ATATATTCCCTCTTTCATTCTCCTTCTTCGCTTTTCTTTGTTTCTCATATTC AAGACAT
CCTCAATTCCAAATCTTAAACCCTAAATTTACAGACACAATCGAGATCACCTGAAAAAAG
AGGTTTAAAGATTTTAGCAAAGATGGCGAATTCAGGAAATTATGGAAAGAGGCCCTTTCG
AGGCGATGAATCGGATGAAAAGAAAGAACCGATGATGATGAGAACATATTCCTTTCTT
CTCTGCCCCGATCCCAATATGACATGCGTGCCATGGTCTCAGCCTTGACTCAAGTCATGG
AAACCAAAGCAGCTCTCATGATAATAACCAACATCAACCTGTTGTGTATAATCAACAAGA
TCCTAACCACCGGCTCCTCCAACCTCAAGATCAAGGGCTATTGAGGAAGAGGCATATAG
AGGGGTAAAGACAACGACCATGGGGAAGTGGGCAGCTGAAATTCGGGATCCGCAAAAGGC
AGCACGGGTGTGGCTCGGGACATTTGAGACTGCTGAAGCTGCGGCTTTAGCTTATGATAA
CGCAGCTCTTAAGTTCAAAGGAAGCAAAGCCAACTCAATTTCCCTGAGAGAGCTCAACT
AGCAAGTAACACTAGTACAACCTACCGGTCCACCAAATATTATCTTCTAATAATCAAAT
TTACTACTCAAATCCGAGACTAATCCGCAAACCATACCTTATTTTAACCAATACTACTA
TAACCAATATCTTTCATCAAGGGGGGAATAGTAACGATGCATTAAGTTATAGCTTGGCCGG
TGGAGAAACCGGAGGCTCAATGTATAATCATCAGACGTTATCTACTACAAATCTTTCATC
TTCTGGTGGATCTTCAAGGCAACAAGATGATGAACAAGATTACGCCAGATATTTGCGTTT
TGGGGATTCTTCACCTCCTAATTCTGTTTTTGGAGATCTTCAATAAACTGATAATAAAGG
ATTTGGGTCACTTGTTATGAGGGGATCATATGTTTTCTAA

>G2509 Amino Acid Sequence (domain in aa coordinates: 89-156)

MANSGNYGKRPFRGDESDEKKEADDDENIFPFFSARSQYDMRAMVSALTQVIGNQSSSHD
NNQHQPVVYNQQDPNPPAPPTQDQGLLRKRYRGVRQRPWGKWA AEIRD PQKAARVWLGT
FETAEEAALAYDNAALKFKGSKAKLNFPERAQLASNTSTTTGPPNYYSNNQIYYSNPQT
NPQTIPIFYFNQYYNQLYHQGNSNDALSYSLAGGETGGS MYNHQTLSTTNSSSSGGSSRQ
QDDEQDYARYLRFGDSSPPNSGF*

>G390 (1..2526)

ATGATGGCTCATCACTCCATGGACGATAGAGACTCTCCTGATAAAGGATTTGATTCCGGC
AAGTACGTTAGATACAGCCGGAACAAGTTGAAGCTCTTGAGAGAGTTTATGCTGAGTGT
CCTAAACCTAGCTCTCTGAGAAGACAACAGCTTATTCGTGAATGTCCCATTCTCTGTAAC
ATCGAGCCTCGACAGATCAAAGTTTGGTTCCAGAATCGCAGATGTCGAGAGAAGCAGAGG
AAAGAGTCAGCTCGTCTTCAGACAGTGAACAGGAAGCTGAGTGCTATGAACAAGCTTTTG
ATGGAAGAGAATGATCGTTTGCAGAAGCAAGTCTCCAACCTTGCTTATGAGAATGGATTG
ATGAAACATCGAATCCACACTGCTTCTGGGACGACCACAGACAACAGCTGTGAGTCTGTG
GTCGTGAGTGGTCAGCAACGTCAGCAGCAAACCCAACACATCAGCATCCTCAGCGTGAT
GTTAACAACCCAGCTAATCTTCTCTCGATTGCGGAGGAGACCTTGGCGGAGTTCCTTTGC
AAGGCTACAGGAACCTGTGCTGACTGGGTCCAGATGATTGGGATGAAGCCTGGTCCGGAT
TCTATTGGTATCGTAGCTGTTTACGCAACTGCAGTGAATAGCAGCACGTGCCTGTGGC
CTCGTGAGTTTAGAACCCATGAAGGTGCTGAAATCCTCAAAGATCGTCCATCTTGGTTC
CGTGACTGTGATGTGTGAGACTCTGAATGTTATACCCACTGGAAATGGTGGTACTATC
GAGCTTGTCAACACTCAGATTTATGCTCCTACAACATTAGCAGCAGCTCGTGACTTTTGG
ACGCTGAGATATAGTACAAGTCTAGAAGATGGAAGCTATGTGGTCTGTGAGAGATCACTC
ACTTCTGCAACTGGTGGCCCCAATGGTCCACTTTCTTCAAGCTTCGTGAGAGCCAAATG
CTGTCAAGCGGGTTTCTTATCCGTCCTTGTGATGGTGGTGGTTCATTATTCACATCGTT
GATCATGTGGACTTGGATGTCTCAAGTGTTCCTGAAGTCTCAGGCCTCTTTATGAGTCT
TCCAAATCCTTGCTCAAAAATGACTGTGCTGCTCTGAGACATGTGCGCCAAATTGCT
CAAGAGACTAGTGGAGAAGTCCAGTATAGTGGTGGACGCCAGCCTGCAGTTTTTAAGGACT
TTCAGCCAGAGACTCTGCCGGGGTTTCAATGATGCTGTAAATGGTTTTGTGCGATGATGGA
TGGTCTCCAATGAGTAGTGATGGAGGAGAGGATATTACGATCATGATTAACTCTTCTCT
GCTAAATTTGCTGGCTCCCAATACGGTAGCTCATTTCTTCCAAGTTTTGGAAGTGGTGT
CTCTGTGCCAAAGCTTCTATGCTGTTGCGAATGTTCCACCCCTTGATTGATTCCGTTG
CTGAGAGAACACCGAGCTGAATGGGCAGACTATGGTGTGCGATGCTTCTGCTGCATCT
CTCAGAGCAACTCCATATGCTGTTCCATGCGTCAGAACCGGTGGGTTCCCGAGTAACCAA
GTCATTCTTCTCTCGCACAGACACTCGAACATGAAGAGTTTCTCGAAGTGGTTAGACTT

GGAGGTCATGCTTACTCACCTGAAGACATGGGCTTATCCCGGGATATGTATTTACTGCAG
CTTTGTAGCGGCGTTGATGAAAAATGTGGTTGGAGGTTGTGCTCAGCTTGTCTTTGCCCA
ATCGATGAATCATTGCTGATGATGCACCTTTGCTTCCTTCTGGTTTCCGTGTCATACCA
CTCGACCAAAAAACAAATCCGAATGATCATCAATCTGCAAGTCGAACACGGGATCTAGCA
TCGTCCTTAGATGGTTCCACCAAAACCGATTCCGAAACAACTCTAGATTGGTCTTAACA
ATAGCCTTCCAGTTCACGTTTGATAACCATTCAGAGACAATGTTGCTACAATGGCGAGA
CAGTATGTGAGGAACGTTGTTGGTTCGATTTCAGAGAGTGGCTCTAGCCATTACGCCTCGT
CCTGGCTCAATGCAACTTCCCACTTCCCTGAAGCTCTCACTCTTGTCCGTTGGATCACC
CGTAGTTACAGTATTCATACAGGTGCAGATCTGTTTGGAGCTGATTCTCAGTCCTGTGGA
GGAGACACATTGCTTAAGCAACTCTGGGACCATAGTGATGCCATATTGTGCTGCTCCCTG
AAAATAATGCCTCACCGGTATTACATTTGCAAACCAAGCTGGTTTAGACATGCTTGAA
ACTACACTTGTGGCACTTCAGGATATAATGCTCGACAAAACACTTGATGACTCTGGTCGT
AGAGCTCTTTGCTCCGAGTTCGCCAAGATCATGCAGCAGGGATATGCGAATCTTCCGCA
GGAATATGTGTGTCGAGCATGGGCAGACCGGTTTTCGTATGAGCAAGCGACGGTGTGGAAA
GTTGTTGATGACAACGAATCAAACCACTGCTTGGCTTTTACCCTCGTTAGTTGGTCGTTT
GTTTGA

>G390 Amino Acid Sequence (domain in AA coordinates: 18-81)
MMAHHSMDRSDSPDKGFDGSGKYVRYTPEQVEALERVYAECPKPSSLRRQQLIRECPILCN
IEPRQIKVWFQNRRCREKQRKESARLQTVNRKLSAMNKLMEENDRLQKQVSNLVYENG
MKHRIHTASGTTTNDSCSESVVSGQQRQQNPHTQHPRQDVNNPANLLSTABETLAEFLC
KATGTAVDWVQMIGMKPGPDSIGIVAVSRNCSGIAARACGLVSLPEPMKVAEILKDRPSWF
RDCRCVETLNVPTGNGGTIELVNTQIYAPTTLAAARDFWTLRYSTSLDGSYVVCERSL
TSATGGPNGLSSSFVRAMKLSGFLIRPCDGGGSI IHIVDHVDLDVSSVPEVLRPLYES
SKILAQKMTVAALRHVRQIAQETSGEVQYSGGRQPAVLRTFSQRLCRGFNDVNGFVDDG
WSPMSSDGGEDITIMINSSSAKFAGSQYSSFLPSFGSGVLCASMLLQNVPLVLIRF
LREHRAEWADYGVDAISAASLRATPYAVPCVRTGGFPNSQVILPLAQTLHEEFLEVRL
GGHAYSPEDMGLSRDMLYLQLCSGVDENVVGGCAQLVFAPIDESFADDAPLLPSGFRVIP
LDQKTNPNHDQSASRTDLASSLDGSTKTDSETNSRLVLTIAFQFTFDNHSRDNVATMAR
QYVRNVVGSIQRVVALAITPRPGSMQLPTSPEALTLVRWITRSYSIHTGADLFGADSQSCG
GDTLLKQLWDHSDAILCCSLKTNASPVFTFANQAGLDMLETTLVALQDIMLDKTLDDSGR
RALCSEFAKIMQQGYANLPAGICVSSMGRPVSYEQATVWKVDDNESNHCLAFTLVSWSF
V*

>G391 (1..2559)

ATGATGATGGTCCATTTCGATGAGCAGAGATATGATGAACAGAGAGTCGCCGGATAAAGGG
TTAGATTCCGGCAAGTATGTGAGGTACACGCCGGAGCAAGTGGAAGCTCTCGAGAGAGTT
TACACTGAGTGTCTTAAGCCAAGTTCTCTAAGAAGACAACAACCTCATACGTGAATGTCCG
ATTCTCTCTAACATCGAGCCTAAGCAGATCAAAGTTTGGTTTCAGAACCGCAGATGTCTGT
GAGAAGCAGAGGAAAGAAGCTGCTCGTCTTCAAACAGTGAACAGAAAACCTCAATGCCATG
AACAAACTCTTGATGGAAGAGAATGATCGTTTGCAGAAGCAAGTTTCTAACTTGGTCTAT
GAGAATGGCCACATGAAACATCAACTTCACACTGCTTCTGGGACGACCACAGACAACAGC
TGTGAGTCTGTGGTCGTGAGTGGTCAGCAACATCAACAGCAAAACCCAAATCCTCAGCAT
CAGCAACGTGATGTAAACAACCCAGCAGGACTCCTTTCTATAGCAGAGGAGGCCCTAGCA
GAGTTCCTTTCCAAGGCTACAGGAACCTGCTGTTGACTGGGTTTCAGATGATTGGGATGAAG
CCTGGTCCGGATTCTATTGGCATACTCGCTATTTCCGCGCAACTGCAGCGGAATTGCAGCA
CGTGCTGCGGCCTCGTGAGTTTGAACCCATGAAGGTTGCTGAAATTCTCAAAGATCGT
CCATCTTGGCTCCGAGATTGTGCAAGTGTGGATACTCTGAGTGTGATACCTGCTGGAAAC
GGTGGGACGATCGAGCTTATTTACACGCAGATGTATGCTCCTACGACTTTAGCAGCAGCT
CGTGACTTTTGGACGCTGAGATATAGCACATGTTTGAAGATGGAAGCTATGTGGTTTGT
GAAAGGTCGCTTACTTCTGCAACTGGTGGCCCCACTGGGCCACCTTCTTCAAACCTTGTG
AGAGCTGAAATGAAACCAAGCGGGTTTCTCATCCGTCCTTGCGATGGTGGTGGTTCCATT
CTCCACATTGTTGATCATGTTGATCTGGATGCCTGGAGTGTCCCTGAAGTCATGAGGCCT
CTCTATGAATCATCGAAGATTCTTGCTCAGAAAATGACTGTTGCTGCTTTGAGACATGTA
AGACAAATTGCAAGAAACAAGTGGAAGTTTCAATGATGGTGGAGGGCGCAACCTGCG
GTTTTAAGAACCTTCAGTCAAAGACTCTGTCGGGGTTTCAATGATGCTGTTAATGGTTTT
GTGGATGATGATGGTACCAATGGGTAGCGATGGTGCAGAGGATGTTACTGTAATGATA
AAGTTGTCCTCGGGAAGTTTGGTGGTCTCAGTACGGTAATTCAATTCCTTCCAAGCTTT
GGTAGTGGCGTGCTTTGTGCCAAGGCATCTATGTTGCTTCAGAACGTTCCACCCGCTGTG

CTGGTTCGATTCCCTTAGAGAACACCGATCTGAATGGGCTGATTATGGCGTGGATGCTTAT
GCTGCTGCATCGCTCAGAGCAAGTCCTTTTGTGCTTCCTTGTGCTAGAGCTGGGGGGTTC
CCAAGTAACCAAGTCATTCTTCCTCTTGCAGACAGTTGAACATGAAGAGTCACTTGAG
GTGGTTAGACTTGAAGGTCACGCTTACTCACCCGAAGACATGGGTTAGCTCGGGATATG
TATTTGCTACAGCTTTGTAGCGGTGTGTATGAAAATGTGGTTGGAGGTGTGCACAGCTT
GTATTTGCCCCATCGATGAATCATTTGCTGATGATGCACCTTTGCTTCCTTCCGGTTTC
CGCATACACTCTTGAACAGAAATCTACTCCGAACGGTGCATCTGCAAACCGTACCTTG
GATTTAGCCTCAGCTTTAGAAGGATCCACACGTCAAGCTGGTGAAGCCGACCCAAATGGC
TGTAACCTTAGGTGCGTACTAACCATAGCATTCCAGTTCACATTTGATAACCATTCAAGA
GACAGTGTGCTTCAATGGCACGTACGTACGTGCGAAGCATAGTAGGATCGATTAGAGG
GTTGCTCTAGCCATTGCTCCTCGTCCGCTCCAATATCAGTCCAATATCTGTTCCCACT
TCCCCTGAAGCTCTCACTCTGGTCCGTGGATCTCCCGGAGTTACAGCCTTCACACTGGT
GCAGATCTCTTTGGATCTGATTCTCAAACAGTGGTGACAGTGTGCTGCATCAACTCTGG
AATCACTCTGATGCAATCTTGTGCTGCTCCCTCAAACAAACGCTTCACCGGTTTTTACA
TTCGCAAACAAACCGGTTTAGACATGCTGGAACGACTCTTGTAGCCCTTCAAGACATA
ATGCTAGACAAGACCTTGACGAACCTGGTCGTAAAGCTCTTGTCTGAGTTCCCAAG
ATCATGCAACAGGGCTATGCTCATCTGCCGCGAGGAGTATGTGCGTCAAGCATGGGAAGG
ATGGTATCTTACGAGCAGGCAACGGTGTGGAAGTTCTTGAAGACGATGAATCAAACCAC
TGCTTAGCTTTCATGTTCTGTAATTGGTCTGTTCTGTTTGA

>G391 Amino Acid Sequence (domain in AA coordinates: 25-85)
MMMVHMSRDMNRESPPDKGLDSGKYVRYTPEQVEALERVYTECPKSSLRQQLIRECP
ILSNIEPKQIKVWFQNRRCREKQKBAARLQTVNRKLNAMNKLMEENDRLQKQVSNLVY
ENGHMKHQLHTASGTTTNDNSCESVVVSGQQHQQNPNPQHQQRDANNPAGLLSIAEEALA
EFLSKATGTAVDWVQMIGMKPGPDSIGIVAISRNCSGIAARACGLVSLPEMKVAEILKDR
PSWLRCRSDVTLVSVIPAGNGGTIELIYTQMYAPTTLAAARDFWTLRYSTCLEDSYVVC
ERSLTSATGGTGPSSNFVRAEMKPSGFLIRPCDGGGSILHIVDHVDLDAWSVPEVMRP
LYESSKILAQKMTVAALRHVRQIAQETSSEVQYGGGRQPAVLRTFSQRLCRGFNDVNGF
VDDGWSPPMGSDGAEDVTVMINLSPGKFGGSQYGNFSLPSFGSVLCAKASMLLQNVPPAV
LVRFLEHRSEWADYGVDAAYAAASLRASPFVPCARAGGFPSNQVILPLAQTVHEESLE
VVRLEGHAYSPEMDGLARDMYLLQLCSGVDENVVGGCAQLVFAPIDESFADDAPLLPSGF
RIIPLEQKSTPNGASANRTLDLASALEGSTRQAGEADPNCGNFRSVLTIAFQFTFDNHSR
DSVASMARQYVRSIVGSIQNRVALAIAPRPGSNISPISVPTSPEALTLVRWISRSYSLHTG
ADLFGSDSQTSGDTLLHLQWNSDAILCCSLKTNASPVFTFANQTGLDMLETTLVALQDI
MLDKTLDEPGRKALCSEFPKIMQQGYAHLPAVCASSMGRMVSVEQATVWKVLEDDSNH
CLAFMFVNWSFV*

>G438 (188..2716)

CGGGGTACCAAGCCACGACCGTAGAATCTTCTTTGTCTGAAAAGAATTACAATTTACG
TTTCTCTTACGATACGACGGACTTTCCGAAGAAATTAATTTAAAGAGAAAAGAAGAA
GCCAAAGAAGAAGAAGTCTAGAAGAAACAGTAAAGTTTGAGACTTTTTTTGAGGGTCG
AGCTAAATGGAGATGCGCGTGGCTAACCACCGTGAGAGAAGCAGTGACAGTATGAATAG
ACATTTAGATAGTAGCGGTAAGTACGTTAGGTACACAGCTGAGCAAGTCGAGGCTCTGA
GCGTGTCTACGCTGAGTGTCTAAGCCTAGCTCTCTCCGTCGACAACAATTGATCCGTGA
ATGTTCCATTTTGGCCAATATTGAGCCTAAGCAGATCAAAGTCTGGTTTCAGAACCGCAG
GTGTCGAGATAAGCAGAGGAAAGAGGCGTCGAGGCTCCAGAGCGTAAACCGGAAGCTCTC
TGCGATGAATAAATGTTGATGGAGGAGAATGATAGGTTGCAGAAGCAGGTTTCTCAGCT
TGTCTCGGAAATGGATATATGAAACAGCAGCTAACTACTGTTGTTAACGATCCAAGCTG
TGAATCTGTGGTCACAACTCCTCAGCATTCGCTTAGAGATGCGAATAGTCTCTGCTGGATT
GCTCTCAATCGCAGAGGAGACTTTGGCAGAGTTCTTATCCAAGGCTACAGGAAGTGTGT
TGATTGGGTTTCAGATGCCTGGGATGAAGCCTGGTCCGGATTTCGGTTGGCATCTTTGCCAT
TTCGCAAAGATGCAATGGAGTGGCAGCTCGAGCCTGTGGTCTTGTGTAGCTTAGAACCTAT
GAAGATTGCGAGATCCTCAAAGATCGGCCATCTTGGTTCCGTGACTGTAGGAGCCTTGA
AGTTTTCACTATGTTCCCGGCTGGTAATGGTGGCACAATCGAGCTTGTATATGCGAGAC
GTATGCACCAACGACTCTGGCTCCTGCCCGGATTCTTGACCCCTGAGATACACAACGAG
CCTCGACAATGGGAGTTTTGTGGTTTGTGAGAGGTCGCTATCTGGCTCTGGAGCTGGGCC
TAATGCTGCTTCAGCTTCTCAGTTTGTGAGAGCAGAAATGCTTTCTAGTGGGTATTTAAT
AAGGCCTTGTGATGGTGGTGGTCTATTATTACATTGTGATCACCTTAATCTTGAGGC
TTGGAGTGTCCGGATGCTTCCGACCCCTTTATGAGTCATCAAAGTCGTGACAAAA

AATGACCATTTCCGCGTTGCGGTATATCAGGCAATTAGCCCAAGAGTCTAATGGTGAAGT
 AGTGTATGGATTAGGAAGGCAGCCTGCTGTTCTTAGAACCTTTAGCCAAAGATTAAGCAG
 GGGCTTCAATGATGCGGTTAATGGGTTTGGTGACGACGGGTGGTCTACGATGCATTTGTGA
 TGGAGCGGAAGATATTATCGTTGCTATTAACCTCTACAAAGCATTTGAATAATATTTCTAA
 TTCTCTTTTCGTTCCCTTGGAGGCGTGCTCTGTGCCAAGGCTTCAATGCTTCTCCAAAATGT
 TCCTCCTGCGGTTTTTGATCCGGTTCCTTAGAGAGCATCGATCTGAGTGGGCTGATTTCAA
 TGTTGATGCATATTCCGCTGCTACACTTAAAGCTGGTAGCTTTGCTTATCCGGGAATGAG
 ACCAACAAGATTCCTGGGAGTCAGATCATAATGCCACTAGGACATACAATTGAACACGA
 AGAAATGCTAGAAGTTGTTAGACTGGAAGGTCATTCTCTTGCTCAAGAAGATGCATTTAT
 GTCACGGGATGTCCATCTCCTTCAGATTTGTACCGGGATTGACGAGAATGCCGTTGGAGC
 TTGTTCTGAACTGATATTGCTCCGATTAATGAGATGTTCCCGGATGATGCTCCACTTGT
 TCCCTCTGGATTCCGAGTCATACCCGTTGATGCTAAAACGGGAGATGTACAAGATCTGTT
 AACCGCTAATCACCCTACACTAGACTTAACCTTAGCCTTGAAGTCGGTCCATCACCTGA
 GAATGCTTCTGGAAGCTCTTTTCTAGCTCAAGCTCGAGATGTATTCTCACTATCGCGTT
 TCAATTCCTTTTGAACCAACTTGCAAGAAAATGTTGCTGGTATGGCTTGTCAGTATGT
 GAGGAGCGTGATCTCATCAGTTCAACGTGTTGCAATGGCGATCTCACCGTCTGGGATAAG
 CCCGAGTCTGGGCTCCAAATTGTCCCAGGATCTCTGAAGCTGTTACTCTTGCTCAGTG
 GATCTCTCAAAGTTACAGTCATCACTTAGGCTCGGAGTTGCTGACGATTGATTCACCTTGG
 AAGCGACGACTCGGTACTAAAACCTTATGGGATCACCAAGATGCCATCCTGTGTTGCTC
 ATTAAGCCACAGCCAGTGTTCATGTTTGCGAACCAAGCTGGTCTAGACATGCTAGAGAC
 AACACTTGTAGCCTTACAGATATAACACTCGAAAAGATATTGATGAATCGGGTCGTAA
 GGCTATCTGTTGCGACTTCGCCAAGCTAATGCAACAGGGATTGCTTGCTTGCCTTCAGG
 AATCTGTGTGTCAACGATGGGAAGACATGTGAGTTATGAACAAGCTGTTGCTTGGAAAGT
 GTTTGCTGCATCTGAAGAAAACAACAATCTGCATTGTCTTGCTTCTCCTTTGTA
 CTGGTCTTTTGTGTGATTGATGACAGAAAAGACTAATTTAAATTTACGTTAGAGAAC
 TCAAATTTTGGTTGTGTTTAGGTGTCTCTGTTTTGTTTTTAAATTTATTTTGATCAA

A

>G438 Amino Acid Sequence (domain in AA coordinates: 22-85)
 MEMAVANHRRSSDSMNRLDSSGKYVRYTAEQVEALERVYAECPKPSSLRRQQLIRECS
 ILANIEPKQIKVWFQNRRCRDKQRKEASRLQSVNRKLSAMNKLMEENDRLQKQVSQLVC
 ENGYMKQLTTFVNDPSCSVVTTTQHSRLDANSFAGLLSIAEETLAEFLSKATGTAVDW
 VQMPGMPKPGPDSVGIFAIQSQRNGVAARACGLVSLPEMKIAEILKDRPSWFRDCRSLEVF
 TMFPAGNGGTIELVYMQTYAPTTLAPARDFWTLRYTSLDNGSFVVCERSLSGSGAGPNA
 ASASQFVRAELSSGYLIRPCDGGGSIHIVDHLNLEAWSVPDVLRLPLYESSKVVAQKMT
 ISALRYIRQLAQESNGEVVYGLGRQPAVLRTFSQRLSRGFNDVNGFGDDGWSTMHCDGA
 EDIIVAINSTKHLNNISNLSFLGGVLCASMLLQNVPPAVLIRFLREHRSEWADFNV
 AYSAAATLKAGSFAYPGMRPTRFTGSQIIMPLGHTIEHEEMLEVVRLEGHSLAQEDAFMSR
 DVHLLQICTGIDENAVGACSELIIFAPINEMFPDDAPLVP SGFRVIPVDAKTGDVQDLTA
 NHRTLDLTSSLEVGPSPENASGNSFSSSSSRILTIAFQFPFENNLOENVAGMACQYVRS
 VISSVQVRVMAISPSGISPSLGSKLSPGSPEAVTLAQWISQSYSHHLGSELLTIDSLGSD
 DSVLKLWDHQDAILCCSLKPQPVFMFANQAGLDMLETTLVALQDITLEKIFDESGRKAI
 CSDFAKLMQQGFACLPSCICVSTMGHVSIEYQAVAWKVFAASEENNNNLHCLAFS FVNWS
 FV*

>G47 (38..472)

CTTCTTCTTCACATCGATCATCATACAACAACAAAAATGGATTACAGAGAATCCACCGG
 TGAAAGTCAGTCAAAGTACAAAGGAATCCGTCGTCGGAATGGGGCAAATGGGTATCAGA
 GATTAGAGTTCCGGGAACCTGTCACCGTCTCTGGTTAGGTTCACTCTCAACAGCAGAAGG
 TGCCGCGGTAGCACACGACGTTGCTTTCTCTGTTTACACCAACCTGATTCTTTAGAATC
 TCTCAATTTCCCTCATTTCGCTTAATCCTTCACTCGTTTCCAGAACCTTCCGAGATCTAT
 CCAGCAAGCTGCTTCAACGCCGGCATGGCCATTGACGCCGGAATCGTCCACAGTACCAG
 CGTGAACCTCTGGATGCGGAGATACGACGACGTATTACGAGAATGGAGCTGATCAAGTGA
 GCCGTTGAATATTTAGTGTATGATTATCTGGGCGGCCACGATCACGTTGATTTATCTC
 GACGGTCATGATCACGTTTGATCTTCTTTGAGTAAGATTTTGTACCATAATCAAAACAG
 GTGTGGTGTAAAAATCTTACTCAAAACAAGATTAGGTACCAAGAGAAAACAAATCAATGG
 TTGTGAATATACATTATAAGGTTTTGATTAAATGTTTGTTCCTGATTAGTGAAGTTG
 GTCCATTGTATACAAATCTATTCAAGAAACCTAGCGCGAGATCATGTTTCGTGATTGAAG
 ATTGAGATTTTAAAGTATTCGTAATATTTTGTAAATACAAATAAAAAAAAAAAAAAA

AAAAA

>G47 Amino Acid Sequence (domain in AA coordinates: 11-80)
MDYRESTGESQSKYKGI RRRKWKVSEIRVPGTRDRLWLGSFSTAEGAAVAHDVAFFCL
HQPDSLES LNFP HLLNPSLVSR TSPRSIQQAASNAGMAIDAGIVHSTSVNSGCGD TTTY
ENGADQVEPLNISVYDYLGGHDHV*

>G559 (89..1285)

aaagttgctagctttaatttgccaacttactattcttatgtgtaataatcgtttgagg
tcggtgatttgggtgataagtcagtagaaATGgataaggagaaatctccagcacctccttg
tggaggtcttcctcctccatctccatcaggtcgatgctctgcattctcagaagctggtcc
cattggtcatggttcagatgctaatacgatgagtcgatattagccgatgcttgataa
cccacctaagaagattggacatcggcgagctcattctgaaatacttactctccctgatga
tttgagctttgatagtgatcttggtgtggttggaatgctgctgatggagcttcttctc
tgatgagactgaagaagatttgctctctatgtatcttgatagggataagtttaattcttc
tgctacatcttctgcccagttggtgagccatcaggaactgcttggaataatgagacaat
gatgcagacaggcacaggctcaacttccaatcctcagaatacgggttaatagctcttgccga
aaggccaagaatcaggcatcaacatagccaatctatggatggttcaatgaatatcaatga
gatgcttatgtcgggaaatgaagatgattctgctattgatgctaagaagtctatgtctgc
tactaaacttgctgagcttgctctcattgatcctaaacgtgctaagaggatatgggcaa
caggcagtcgcgacgacgatcaaaagaaaggaagacgagatacatatttgagcttgagag
aaaagtacagactttgcaaacagaggctacaactctctcagcccagttgacctcttaca
gagagacacaaatggcttgactggtgaaaacaatgagctgaagctgcggttacaacaat
ggagcagcaggttcacttgaggatgaactaaacgaagcactaaaggaggaaatccagca
tctgaaggtgttgactggccaggttgctccatcagcgttgaactatgggtcggttggtc
aaaccagcagcaattctattccaacaatcagtcgaatgcaacaatcttagctgcaaaaca
gttccagcaacttcagattcattcacagaagcagcaacaacaacaacaacaacaaca
gcaacaccaacagcagcagcagcaacagcaacagtcagtttcaacagcaacagatgca
acagcttatgcagcagcggcttcaacagcaagaacaacaaatggagtaagactcaagcc
ttcacaagcccagaaagagaacTGAggaatatgaatatgtcccacgtaagtgaaggttc
tccttctgaacaattcctttctcattcataaattggttggtcattccatcacttgagcttc
ttggatttttagggtttttagctaacaca

>G559 Amino Acid Sequence (domain in AA coordinates: 203-264)

MDKEKSPAPPCGGLPPPSPSGRCSAFSEAGPIGHGSDANRMSHDISRMLDNPPKKIGHRR
AHSEILTL PDDL SFDSLGVVGN AADGASFSD ETEEDLLSMYLDMDKFNSSATSSAQVGE
PSGTAWKNETMMQTGTGSTSNPQNTVNSLGERPRIRHQHSQSMDSMNINEMLSGNEDD
SAIDAKKSMSATKLAEALALIDPKRAKRIWANRQSAARSKERKTRYIFELERKVQLQTEA
TTL SAQLTLLQRDTNGLTVENNELKLRLQTM EQQVHLQDELNEALKEEI QHLKVL TGQVA
PSALNYGSFSGNQQQFYNNQSMQTILAAKQFQQLQIHSQKQQQQQQQQQHQHQHQHQ
QQYQFQQQQMQQLMQQLQQEQNGVRLKPSQAQKEN*

>G568 (141..995)

GACCGGCTAAAGTCAAGAACCTCTCTCTGAGCTCTCACCCTTTCTCTCTACTCCCTC
TCTGCGTGTAGGATACTACTAGACAATTGACAACCAAGACTAAAGCTGTGTGTTGGTT
CACTTCTGTTCTCTTTTCCAATGTTGTCATCAGCTAAGCATCAGAGAAACCATAGACTCT
CTGCTACAAACAAGAACCAGACTCTCACCAGTCTTCTTCCATTTTCATCCTCATCACCAT
CGTCTTCTTCTTCATCATCATCAACCTCATCATCTCCTTTACCTTCTCAAGACTCTC
AAGCCCAGAAGAGATCTCTTGTCACCATGGAAGAAGTTTGAATGACATCAACCTTGCTT
CCATCCACCACCTAAACCGACACAGCCCTCATCCACAACACAACCACGAGCCAAGGTTCA
GGGGCCAAAACCACCAACCAAAACCTAACTCAATCTTCCAAGATTTTCTCAAAGGAT
CTTTGAACAGGAACAGCACCCACAAGCCAGACCACGGGTTCTGCGCCTAATGGCGATT
CCACCACGGTCACTGTTCTTTACAGCTCTCCTTTTCCACCTCCTGCAACTGTTCTGAGCT
TGAATTCGGCGCTGCTTCGAGTTCTCGATAACCAAGATCCTCTTGTACCTCAAACCT
CTAATCTTCATACCCACCATCACCTCTCAAACGCTCATGCCCTCAACACCTCTTTCGAGG
CTCTGGTTCCATCCAGTTCTTTTGGTAAGAAAAGAGGCCAAGATTCCAATGAAGGTTTCA
GGAAATAGAAGACATAAGCGTATGATCAAGAACAGAGAATCTGCAGCTCGTTCCCGCGCTA
GGAAACAGGCTTATACAAACGAGTTAGAAGTTGAAGTTGCTCACTTGCAGGCAGAAAATG
CAAGACTCAAGAGACAACAAGATCAAAAAATGGCTGCAGCAATTGAGCAACCCAAAAAGA
ACACACTTCAACGGTCTTCCACAGCTCCATTTTGAGAAATCTACAAGTCTTGTCTCTCT
TTTGGGGATTGAGATTGTCTCATGAAGAAGTGAAGAAATGGCAAGTTTGTACCTTTT

TTATTAGCTATAAGTATAACTAAGCCTAAAATTGTAGAACTAAGATATTGTAGGGGAAAA
AAGAAGATGTAAACAAAAGACCCGGAAAGAGAAAAGGATCTTTCAATTTCTTAAGGCAC
AGGAACACCTGTCTGGGTCTCTCTTAATGTTCTGTCTGTTTTCCTATGCAAACCCCTTT
TTCACCTCTGTACTAACTTATACTTGTATTCTTG
>G568 Amino Acid Sequence (domain in AA coordinates: 215-265)
MLSSAKHQNRHRLSATNKNQTLTKVSSISSSSPSSSSSSSSSSPLPSQDSQAQKRSL
VTMEEVWNDINLASIHHLNRHSPHPQHNHEPRFRGQNHNNQNPNSIFQDFLKGSLNQEP
PTSQTTGSAPNGDSTTVTVLYSSPFPPTATVLSLNSGAGFEFLDNQDPLVTSNSNLHTHH
HLSNAHAFNTSFEALVPSSSFGKKRGQDSNEGSGNRRHKRMIKNRESAARSARKQAYTN
ELELEVAHLQAENARLKRQDQKMAAAIQPKKNTLQRSSTAPF*
>G580 (43..747)
CCAAAAACAAAGCATTCTATGCTATTCTGTTCTGTTCTCCAATGTTGTCATCAGCAAAG
CATAATAAGATCAACAACCATAGTGCCTTTTCAATTTCTCTTCATCATCATCATTTATCA
ACATCATCTCCCTAGGCCATAACAAATCTCAAGTCACCATGGAAGAAGTATGGAAAGAA
ATCAACCTTGGTTCACCTTCACTACCATCGGCAACTAAACATTGGTCATGAACCAATGTTA
AAGAACAAAACCTAATAACTCCATCTTTCAAGATTTCTCAACATGCCTCTGAATCAA
CCACCACCACCACCACCACCCTTCTCTTCCACCATTGTCACTGCTCTCTATGGCTCT
CTGCCTCTTCCGCTCTCTGCCACTGTCTCAGCTTAACTCCGGTGTGGATTGAGTTT
CTTGATACCAAGAAAATCTTCTGCTTCTAACCCTCGCTCCTTTGAGGAATCTGCAAAG
TTTGGTTGCTCTGGTAAGAAAAGAGGCCAAGATTCTGATGATACTAGAGGAGACAGAAGG
TATAAGCGTATGATCAAGAACAGAGAATCTGCTGCTCGTTCAAGGGCTAGGAAGCAGGCA
TATACAAACGAACCTTGAGCTTGAAATTGCTCACTTGCAGACAGAGAATGCAAGACTCAAG
ATACAAACAGAGCAGCTGAAAATAGCCGAAGCAACTCAAAACCAAGTAAAGAAAACACTA
CAACGGTCTTCCACAGCTCCATTTTGAGAAAATCTACTATTTCTTTTGGGGGAGTTTC
AAGTGTCTTCTATGAAGATGAGAAAAACAGAAAAGTTGTACATTTTAGCTAAGTTAAA
TTTGTGGTGGTAAGTAATGTAAAAGAAAAGTGTGTGTAGAAGAAAAGTGTCTAGAAAAAG
AAAGCAACTAACTTTCTTCTTCTCTCTGTTTCTCTATCAACTCTTTTGACTTTTGACT
TTTTTCTTCTCTACTTAACTCTATTATTGTAATGCCAAGTCAAGTCCTTATCTAGCTA
GTACATGAGTTTCTGTTTTCACTGGTTAAGCCAT
>G580 Amino Acid Sequence (domain in AA coordinates:162-218)
MLSSAKHNKINNHSFAFSSSSSSSLSTSSSLGHNKSQVTMEEVWKEINLGLSLHYHRQLNI
GHEPMLKNQNPNSIFQDFLNMPLNQPPPPPPSSSTIVTALYGSLLPLPPATVLSLNS
GVGFEFLDTTENLLASNPRSFESAKFGCLGKKRGQDSDDTRGDRRYKRMKNRESAARS
RARKQAYTNELELEIAHLQTENARLKIQQEQLKIAEATQNVKTLQRSSTAPF*
>G615 (197..1252)
TTTTTCTTTTCTTCTTTTTTTTTGCTGGTGTGAGAAATTGTACGCTTACTATCTCTCTCT
CTCTCTGCCAGATTCTCTTTTTTGATGATGTGAAAGTTGTGCTTTTGTTCCTTAAGAAA
AAGGCATATTTTAATACTTGATTCTTGGTCTTGTATTCTTGATTCTTGGTTTTTTTTAG
CTTCTTAAGTTCGGTGATGTCTCTTCCACCAATGACTACAACGATGGTAATAACAATGG
AGTGTAACCTCTCTCTCTTTACCTTTCTTCACTCTCTGGCCATCAAGACATCATTCATAA
TCCCTACAACCATCAGTTAAAGCATCTCCGGGCCATATGGTATCAGCAGTTCCTGAATC
TCTGATCGATTACATGGCGTTTAAGTCAAATAATGTTGTGAATCAACAAGGCTTTGAGTT
TCCTGAGGTGTCAAAGGAAATCAAGAAGGTGGTGAAGAAGGACCGACATAGCAAGATTCA
AACGGCAACAAGGATAGAGACAGGAGGGTTAGGCTTTTTATTGGGATTGCTCGCCAATT
CTTTGATCTTCAGGATATGTTGGGGTTTGATAAAGCTAGTAAACGTTAGACTGGCTGCT
CAAGAAGTCAAGAAAAGCCATCAAGAGGTGCTACAAGCAAAAACCTCAACAATGATGA
TGAAGATTTTGGAAACATTGGAGGCGATGTAGAACAAGAAGAGGAGAAGGAGGAGGATGA
CAATGGCGATAAGAGCTTCGTGTATGGTTTGAGCCCCGGGTACGGTGAAGAAGAAGTGGT
ATGTGAGGCCACGAAGGCAGGGATAAGAAAGAAGAAGAGTGAGTTGAGAAACATCTCATC
AAAGGGGCTAGGAGCCAAAGCTAGAGGAAAAGCAAAGGAGCGAACAAAAGAGATGATGGC
CTATGATAATCCAGAGACTGCCTCTGATATTACACAATCTGAAATCATGGACCCATTCAA
GAGGTCTATAGTCTTCAATGAAGGAGAAGATATGACACACCTTTTCTACAAGGAACCAAT
CGAGGAGTTTGATAATCAAGAATCTATCTTAACCAATATGACTCTACCAACGAAGATGGG
TCAAAGTTACAATCAAAATAATGGGATACTTATGTTGGTAGATCAGAGTTCTAGCAGCAA
CTATAATACATTTCTGCCTCAAAATTTGGATTATAGTTATGATCAAAACCTTTTCATGA
CCAAACCTTATATGTAGTCAACGACAAAAATTTCCCAAAGGTTTCTATAAATCTCGAC
AGTTTTGAAGGACTATGCATGATCAAGTTTAAACATGTAAGCCAATATAGTCCCTTATTC

CTCTGAATGTATACAAAATCTATAGTTATGTATATCTGTTCCCTTTTAAACGTATCTTTAT
TGATCTTCTGTGCTTGATCAAAATTGTCTATTTAAGATTCAAGTTTGTGTAATATTTTAG
CTACAACTTTTAAAGTGGTATTATTGTAACCTTTTGAAGTATATATTTTGAAGATGAATAA
GAACATGTTTATATAAAAA

>G615 Amino Acid Sequence (domain in AA coordinates: 88-147)
MSSSTNDYNDGNMNGVYPLSLYLSSLSGHQDIIHNPYNHQLKASPGHMVSAVPESLIDYM
AFKSNVNVNQGFPEFPEVSKEIKKVVKDRHSKIQTAGQIRDRRVRLFFIGIARQFFDLQD
MLGFDKASKTLDWLLKKSRAIKEVVOAKNLNDDDFGNIGGDVEQEEKEEDDNGDKS
FVYGLSPGYGEEVVEATKAGIRKKKSELNRISSKGLGAKARGKAKERTKEMMAYDNPE
TASDITQSEIMDPFKRSIVFNEGEDMTHLFYKEPIEEFDNQESILTNMTLPTKMGQSYNQ
NNGILMLVDQSSSNYNTFLPQNLDSYDQNPFFHDQTLVYVTDKNFPKGFL*

>G732 (73..588)

AAAAAAACCAAAACATAAAACTCTGTCCTTTTTTGTCTTCTTGTAACTTTTCT
TGTTAAAAATCAATGGCGTCATCTAGCAGCACATACCGGAGCTCAAGCTCTCCGACGGT
GGTAATAATAACCCGTCGGACTCCGTCGTCACCGTCGACGAAACGTAAGAAGATG
TTATCGAACAGAGAATCTGCACGTAGGTCAAGGATGCGTAAACAGAAACGTTGATGAT
CTAACGGCTCAGATCAATCAGCTATCAAACGACAACCGTCAGATCTTGAACAGCCTCACC
GTAACATCTCAGCTTTACATGAAGATCCAAGCCGAGAACTCTGTTCTCACCGCTCAGATG
GAGGAGCTTAGCACCAGACTCCAATCTCTCAACGAGATCGTTGATCTTGTTCATCCAAC
GGTGCAGGATTGGTGTGACAGATCGACGGCTGTGGTTTTGATGATCGTACGGTTGGG
ATCGACGGATATTACGATGATATGAATATGATGAGTAATGTTAATCATTGGGGTGGTTCG
GTTTACACTAACCAACCCATTATGGCTAATGATATCAATATGTTATGATTAATAAAATTA
ATTAAATAAATTAGATGCCCCCTTTTTGTCTTTTATTTTAAATTTAGCCCATTGTTGGT
GTTTTTGGGTTGGTGTGATGATGTAATTATAGTACATGCATCTTTGATTGGTTGGAAGGA
TAAATATAAACTTTATATATATATTTGGGGCATATATATATGAGTTGTACTTTGCATGTAT
TGGTGTGTGTTTTGTTATAATTATATGATTATATATGTTTATGTTAAAAAAA

>G732 Amino Acid Sequence (domain in AA coordinates: 31-91)
MASSSSTYRSSSSSDGNNNPSSDVTVDERKRKRLSNRESARRSRMRKQKHVDDLTAQ
INQLSNDNRQILNSLTVTSQLYMKIQAENSVLTAQMEELSTRQLSLNEIVDLVQSNAGF
GVDQIDGCGFDDRTVGIDGYDDMMMSNVNHWGGSVYTNQPIMANDINMY*

>G988 (1..1338)

ATGCTTACTTCCCTTCAAATCCTCTAGCTCCTCCTCCGAAGATGCCACCGCTACCAACCAC
GAGAATCCTCCTCCTTTGTGCATCGCCTCCTCCTCGGCCGCAACCTCCGCCTCACATCAC
CTCCGTCGTCCTTCTTTTACCGCTGCGAATTCGTCTCCAGTCAAACCTCACCGCCGCT
CAAACCTTACTCTCAATCCTCTCCCTTAACTCTTCTCCTCACGGCGACTCCACCGAGCGA
CTTGATACCTCTTCACTAAAGCCTTGTCCGTACGAATCAACCGTCAGCAACAAGATCAG
ACGGCTGAAACGGTTGCCACGTGGACGACGAACGAATGACGATGAGTAACCTCCACGGTG
TTCAGGAGCAGTGATGCAAGAAGACAGTTCTTGTTCGAACCAAGAACAACCAATTCTGAC
TTCGAGTCTTGTTACTATCTTTGGCTAAACCAACTAACGCCGTTTTATTGCGTTTCGGTCAT
TTAACGGCGAACCAAGCTATCCTCGACGCGACGGAGACAAACGATAACGGAGCTCTACAT
ATACTTGATTTAGATATATCACAAGGACTTCAATGGCCTCCATTGATGCAAGCCCTAGCA
GAGAGGTCATCAAACCCCTAGCAGTCCACCTCCATCTCTCCGCATAACCGGATGCGGTCGA
GATGTAACCGGATTAAACCGAAGTGGAGACCGGTTAACCCGGTTTCGCTGACTCTTTAGGT
CTCCAATTCAGTTTACACGCTAGTGATCGTAGAAGAAGATCTCGCCGGACTTTTGCTA
CAGATCCGATTGTTAGCTCTCTCAGCCGTACAAGGAGAGACCATGCGGTCAATTGTGTT
CACTTCCTCCACAAAATATTTAACGACGATGGAGATATGATCGGTCACTTCTTGTGACGG
ATCAAGAGCTTAACTCTAGAATCGTTACAATGGCAGAGAGAGAAGCTAATCATGGAGAT
CACTCGTTCTTGAATAGATTCTCTGAGGCAGTGGATCATTACATGGCGATCTTTGATTCTG
TTGGAAGCGACGTTGCCGCCAAATAGCCGAGAGAGACTAACCCCTAGAGCAACGGTGGTTC
GGTAAGGAGATTTTGGATGTTGTGGCGCGGAAGAGACGGAGAGAAAGCAAAGACATCGG
AGGTTTGAGATTTGGAAGAGATGATGAAGAGGTTTGGTTTCGTTAACGTTCTTATTGGA
AGCTTTGCTTTGTCTCAAGCTAAGCTTCTTCTAGACTTCATTATCCTTCAGAAGGTTAT
AATCTTCAGTTCCCTTAACAATTCTTGTCTTGGCTGGCAAAATCGTCCCTCTCTCTC
GTTTCGTCGTGGAATGA

>G988 Amino Acid Sequence (domain in AA coordinates: 178-195)
MLTSFKSSSSSEDAATTTTENPPPLCIASSAATSASHHLRLLFTAANFVSQSNFTAA
QNLLSILSLNSSPHGDSTERLVHLFTKALSVRINRQQDQTAETVATWTTNEMTMSNSTV

FTSSVCKEQFLFRITKNNSDFESCYYLWLNQLTPFIRFGHLTANQAILDATETNDNGALH
ILDLDISQGLQWPPLMQALAERSSNPSSPPPSLRITGCGRDVTGLNRTGDRLTRFADSLG
LQFQFHTLVIVEEDLAGLLLQIRLLALS AVQGETIAVNCVHFLHKIFNDGDGMIGHFLSA
IKSLNSRIVTMAEREANHGDSFLNRFSEAVDHYMAIFDSLEATLPPNSRERLTLEQRWF
GKEILDVVA AEETERKQRHRRFEIWEEMMKRFGFVNVPISFALSQAKLLRLHYPSEGY
NLQFLNLSLFLGWQNRPLFSVSSWK*

>G1519 (1..1146)

ATGAGGCTTAATGGGGATTCCGGGTCCGGGTCAGGATGAACCCGGTTCGAGCGGGTTTCAC
GGCGGAATCAGACGATTCCCGTTAGCAGCTCAGCCGGAGATTATGAGAGCTGCTGAGAAA
GACGATCAATACGCTTCTTTTCATCCACGAAGCTTGCCGCGATGCCTTCCGACACCTTTTC
GGTACAAGAATCGCTCTTGCTTACCAGAAGGAGATGAAGCTACTTGGACAGATGCTTTAC
TATGTTCTTACGACAGGTTTCAGGGCAACAACTTTAGGAGAGGAATATTGTGACATTATA
CAGGTTGCAGGGCCTTATGGACTCTCTCTTACACCAGCTAGACGTGCTTTGTTCATATTG
TACCAGACCGCAGTTCCATATATCGCAGAGAGAATTAGCACTCGAGCTGCTACGCAAGCA
GTCACCTTTGATGAGTCTGATGAGTTTTTTGGTGATAGTTCATATCCACTCACCAGAAGATG
ATAGATCTTCCATCTTCACTCAAGTTGAACTTCAACTTCTGTAGTATCTAGGTTAAAC
GATAGACTTATGAGATCGTGCCAGCGAGCTATTCAGCGATGGCCTGTGGTTCTTCTGTGTT
GCCCGCGAAGTCTTACAACCTGGTTTTGCGTGCCAATCTGATGCTCTTCTACTTTGAAGGT
TTTTATTATCATATATCGAAACGTGCATCCGGGGTTCGTTATGTTTTTCATAGGAAAGCAA
CTGAATCAGAGACCTAGATACCAAACTTCTGGGGTTTTCTTCTAATCCAATTGTGCATC
CTTGCTGCTGAGGGCTTGCGTCCGAGTAATTTGTCTATCTACTAGCTCCATTTCAGCAG
GCTTCTATAGGATCTTATCAAACTTCAGGAGGGAGAGGTTTACCTGTTTTAAATGAAGAG
GGGAATTTGATAACTTCGGAAGCTGAAAAGGGAACTGGTCTACCTCCGATTCAACTTCA
ACGAGGCGAGTAGGGAATGCACCTCTCTGCTTAAGCACCCGTCAGCACCCAACGGCCACT
CCTTGTTGTCATGTGTTTTGTTGGAGCTGCATTATGGAATGGTGCAACGAGAAGCAAGAA
TGCCCTCTTTGTGCAACGCCCAATACCCATTCAAGTTTGGTTTGTGTTATCATTCTGAT
TTTTAG

>G1519 Amino Acid Sequence (domain in AA coordinates: 327-364)

MRLNGDSGPGQDEPGSSGFHGGIRRFPLAAQPEIMRAAEKDDQYASFIHEACRDAFRHLF
GTRIALAYQKEMKLLGQMLYYVLTTSQGLTGEYCDIIQVAGPYGLSPTPARRALFIL
YQTAVPYIAERISTRATQAVTFDESEFFGDSHIHSPRMLDLPSSSQVETSTSVVSRIN
DRLMRSWHRAIQRPVPLPVAREVLQVLRLANLMLFYFEGFYHISKRASGVRYVFIGKQ
LNQRPYQILGVFLLIQLCILAEGLRRSNLSSITSSIQQASIGSYQTSGGRLPVLNEE
GNLITSEAEKGNWSTSDSTSEAVGKCTLCSTRQHPTATPCGHVFCWSCIMEWCNEKQE
CPLCRTPNTHSSLVCLYHSDF*

>G374 (1..1359)

ATGGACAACAAAAATGATCAGGATATTGATGTTAGATCAGTGGTTGAAGCTGTTTCCGCC
GATCTTTCTTTGGTGCTCCCTCTATGTGGTTGAGAGCATGTGCATGCGCTGCCAAGAA
AATGGAACAACAGATTTCTATTGACCTTAATTCCTCACTTCAGAAAGGTCTTAATATCT
GCATTTGAATGTCCGCATTGCGGGGAAAGGAATAATGAAGTTCAAGTTTCGAGGCGAGATT
CAACCCCGTGATGCTGTTACAATCTAGAGGTTCTAGCTGGTGATGTGAAGATATTTGAC
CGGCAAGTTGTGAAATCTGAATCAGCCACTATTAAGATTCTTGAAGTGGATTTTGAGATT
CCACCAGAGGCCCAACGTGGAAGTTTGTCTACTGTGGAAGGGATATTAGCACGGGCTGCT
GATGAAGTGAAGTGCCTTCAAGAAAGCAAGAAAGTTGATCTTAAACTGCTGAAGCA
ATAGACCAATTTCTGTCCAACTGAGAGCTTGTGCTAAAGCAGAGACATCCTTCACTTC
ATTTTGGATGATCCTGCTGGAAACAGTTTTCATTGAGAACCACATGCTCCATCACCAGAT
CCCTCTCTAACCATCAAATTTATGAGCGAACACCAGAGCAACAAGCAACACTTGGATAT
GTTGCTAACCCTCTCAGGCTGGACAATCAGAAGGAAGCCTTGGCGCACCTGTGATGACT
TTCCCTTCACTTGGCGAGCATGTACGGAGCCGTGTGAGACACGGATGTTCAAATAGAA
ATCCCGTACTTTCAGGAAGTTATTGTCTATGGCATCTACATGTGACAGTTGTGGCTATCGT
AATTCTGAGTTGAAGCCTGTTGGTGCAATTCCTGAAAAGGGAAAGAAAGATTACTCTCTCT
GTGAGGAACATTACAGACCTTAGCCGAGATGTTATCAAGTCGGACACTGCAGGAGTGATA
ATCCAGAACTTGATCTGGAGCTAGCTGGTGGTACACTTGGTGGAATGGTAACAACAGTT
GAAGGGTTGGTTACACAGATCAGAGAAAGCCTAGCGAGAGTTACGGATTCACTTTTGGT
GATAGTATGGAAGAGAGTAAGTTGAACAAATGGAGAGAATTTGGAGCCAGGCTCACTAAG
CTCCTAAGCTTTGAACAGCCGTGGACATTGATTCTTGATGATGAATTAGCAAATTCCTTT
ATTGCACCAGTAACAGATGATATCAAAGATGACCATCAGCTCACATTTGAAGAGTACGAG

14

MAGFDENVAVMGEWVPRSPSPGTLFSSAIGEEKSSKRVLERELSLNHGQVIGLEEDTSSN
HNKSSQSNVFRGGLSERIAARAGFNAPRLNTEINRTNTDFSIDSNLRSPCLTISSPGLS
PATLLESPVFLSNPLAQPSPTTGKFPFLPGVNGNALSSEKAKDEFFDDIGASFSPHPVSR
SSSSFFQGTTEMMSVDYGNYNRRSSSHQSAEEVKPGSENIENLYGIETDNQNGQNKTS
DVTNTSLETVDHQEEEEQRRGDSMAGGAPAEADGYNWRKYGQKLVKGSEYPRSYKCTN
PNCQVKKKVERSREGHITETIIYKGAHNHLKPPNRRSGMQVDGTEQVEQQQQQRDSAATW
VSCNNTQQQGSNENNVEEGSTRFEYGNQSGSIQAQTGGQYESGDPVVVVDASSTFSNDE
DEDDRGTGHSVSLGYDGGGGGGGEGDESESKRRKLEAFAAEMSGSTRAIREPRVVVQTT
SDVDILDDGYRWRKYGQKVVKGPNPRSYKCTAPGCTVRKHVERASHDLKSVITTYEGK
HNHNDVPAARNSSHHGGGDSGNGNSGGSAAVSHHYHNGHHSEPPRGRFDRQVTTNNQSPFS
RPFSPQPHLGPSPGFSFGLGQTGLVNLSPGLAYGQKMPGLPHPYMTQPVGMSEAMMQR
GMEPKVEPVSDSGQSVYNQIMSRLPQI*

>G1000 (1..954)

ATGGGAAGACCTCCTTGTTGTGACAAGTCCAATGTCAAGAAAGGTCTCTGGACCGAGGAA
GAAGACGCTAAGATCCTTGCTTATGTTGCTATCCATGGTGTAGGAACTGGAGCTTGATC
CCCCAAAAGCAGGTCTGAATCGATGTGAAAGAGCTGTAGACTAAGATGGACTAATTAC
TTAAGACCTGACCTTAAACATGACAGCTTCTTACCCAAGAAGAAGAGCTTATCATTGAG
TGTCATAGAGCCATTGGCAGCAGGTGGTCTTCCATTGCACGAAAGCTTCCAGGAAGAACG
GATAATGATGTGAAGAATCACTGGAACACAAAGCTGAAGAAGAAGCTGATGAAAATGGGG
ATAGACCCGGTGACTCATAAACCGGTTTCTCAACTCCTTGACAGAACTTACAGAAACATTAGC
GGCCATGGAATGCATCCTTCAAAACAGAACCATCTAACAACCTCTATACTCACACAATCC
AACTCAGCTTGGGAAATGATGAGAAACACAACAACAACCATGAGAGTTATTACACCAAC
TCTCCAATGATGTTTACAAATTCCTCTGAGTACCAAACTACTCCATTTTATTCTATAGC
CATCCAAATCATCTGCTCAATGGAACCATCTTTCATGCTCTTCTCATCATCTTCTACT
AGTATCACTCAGCCAAACCAAGTACCTCAAACACCGGTTACTTAACCTTCTACTGGAGCGAT
TTCCTTCTCTCGGACCCGGTTCCCTCAAGTAGTGGGATCCTCAGCTACTAGCGACCTCACT
TTTACGAGAAGCAACATCAATTTCAACATCGAAGCCGAATACATCTCTCAAACATCGAT
TCAAAGGCCTCGGGAACATGTCTTCCGCGAGTTCTTCTGTTGACGAAATACTAGATAAA
GACCAAGAGATGTTGTACAGTTTCTCAACTCTTGAATGATTTCGATTATTAG
>G1000 Amino Acid Sequence (domain in AA coordinates: 14-117)
MGRPPCCDKSNVKKGLWTEEDAKILAYVAIHGVGNWSLIPKKAGLNRCGKSCRLRWNY
LRPDLKHDSPSTQEEELIIECHRAIGSRWSSIARKLPGRDNDVKNHWNTKLKKLKMKG
IDPVTHKPVSQLLAEFRNISGHNASFKTEPSNNSILTQNSAWEMMRNTTTNHESYYTN
SPMMFTNSSEYQTPPFHFYSHPNHLLNGTSSSSSSSSTSITQPNQVPQTPVTNFYWS
FLSDPVPQVVGSSATSDLTFTQNEHHFNIEAEYISQNIIDSKASGTCHSASSFVDEILDK
DQEMLSQFPQLLNDY*

>G1067 (436..1371)

TCTCAAGCTTCTCTCTCTTTTCCCATAGCACATCAGAATCGCTAAATACGACTCCT
ATGCAAAGAAGAGCTACTTCTTCTCTTGCCTAATTAATCTACCTAAGGGTTTCC
TCTTACCTTTCATGAGAGAGATCATTTAACATAAGTCACCTTTTTATATCTTTTGCTTC
GTCTTTAATTTAGTCTCTGTTCTTGGTCTGTTTCTATATTTTGTGGCTTGCCTAACCGAT
CACACCTTAATGCTTTAGCTATTGTTTCTCAAAATCATGAGTTTGTGACTTCTCGATCTG
AGTTTCTTTTTCTCTCTTTACGCTCTTCTTCACTAGCTACCAATATATGAACGAGCAG
GATCAAGAATCGAGAAATTGATTTGAGCTGGCGAATAAGCAGTGGTGGGATAGGGAATTA
GTAGATGCGGCGCGGATGGAAGGCGGTTACGAGCAAGGCGGTGGAGCTTCTAGATACTTC
CATAACCTCTTTAGACCGGAGATTCAACCAACAGCTTCAACCGCAGGGCGGGATCAAT
CTTATCGACGACATCATCATCAGCACCAGCAACATCAACAACAACAACAACCGTCGGAT
GATTCAAGAGAATCTGACCATTTCAAACAAGATCATCATCAACAGGGTCGACCCGATCA
GACCCGAATACATCAAGCTCAGCACCAGGAAAACGTCACGTTGACGTCACACGATCT
AAGAACAAGCCAAAGCCACCGATCATAGTAACTCGTGATAGCCCCAACGCGCTTAGATCT
CACGTTCTTGAAGTATCTCTGGAGCTGACATAGTTGAGAGTGTTCACGTCAGCTAGG
AGGAGAGGAGAGGCGTCTCCGTTTTAGGAGGAAACGGCACCCTATCTAACGTCACCTC
CGTCAGCCAGTCACTCTGGAAATGGCGGTGGTGTGTCGGAGGAGGAGGAGTGTGACT
TTACATGGAAGGTTTGAGATCTTTTCGCTAACGGGGACTGTTTTGCCACCTCCTGCACCG
CCTGGTGCCTGGTGGTTTCTATATTTTAGCCGGAGGGCAAGGTCAGGTGGTGGGAGGA
AGCGTTGTGGCTCCCTTATTGCATCAGCTCCGTTATACTAATGGCGGCTTCGTTCTCA
AATGCGGTTTTTCGAGAGACTACCGATTGAGGAGGAGGAAGAAGAGGTGGTGGTGGCGGA

GGAGGAGGAGGAGGAGGGCCACCGCAGATGCAACAAGCTCCATCAGCATCTCCGCCGTCT
GGAGTGACCGGTGAGGGACAGTTAGGAGGTAATGTGGGTGGTTATGGGTTTTCTGGTGAT
CCTCATTTGCTTGGATGGGGAGCTGGAACACCTTCAAGACCACCTTTTAATTGAATTTT
AATGTCCGGAAATTTATGTGTTTTATCATCTTGAGGAGTCGTCTTTCCTTTGGGATATT
TGGTGTTTAATGTTTAGTTGATATGCATATTTT

>G1067 Amino Acid Sequence (domain in AA coordinates: 86-93)
MEGGEYQGGGASRYFHNLFPEIHHQQLQPQGGINLIDQHHHQHQQHQQQQPSDDSRES
DHSNKDHHQQRPSDPTNTSSAPGKRPRGRPPGSKNKAKPPIIVTRDSPNALRSHVLEV
SPGADIVESVSTYARRRGRGVSVLGGNGTVSNVTLRQPVTPGNGGGVSGGGVVTLHGRF
EILSLTGTVLPPPAPPGAGGLSIFLAGGQGVVGGSVVAPLIASAPVILMAASFSSNAVFE
RLPIEEEEEEGGGGGGGGGGPPQMQQAPSASPPSGVTVGQQLGGNVGGYGFSGDPHLLG
WGAGTPSRPPF*

>G1075 (19..876)
TTTGTGTTTGGTGCTGGCATGGCTGGTCTCGATCTAGGCACAACCTTCTCGCTACGTCCAC
AACGTCGATGGTGGCGGCGGGACAGTTCACCACCGACAACCACCACGAAGATGACGGT
GGCGCTGGAGGAAACCACCATCATCACCATCATAATCACCATCAAGGTTTAGAT
TTAATAGCTTCTAATGATAACTCTGGACTAGGCGGCGGTGGAGGAGGAGGGAGCGGTGAC
CTCGTCATGCGTCGGCCACGTGGCCGTCAGCTGGATCGAAGAACAACCGAAGCCGCCG
GTGATTGTACGCGCGAGAGCGCAAAACACTCTTAGGGCTCACATTCTGAAGTTGGAAGT
GGCTGCGACGTTTTCGAATGTATCTCCACTTACGCTCGTCGGAGACAGCGCGGATTTGC
GTTTATCCGGGACGGGAACCGTCACTAACGTCAGCATCCGTCAGCCTACGGCGGCCGGA
GCTGTTGTGACTCTGCGGGGTACTTTTGAGATTCTTTCCCTCTCCGGATCTTTTCTTCCG
CCACCTGCTCCTCCAGGGGCGACTAGCTTGACGATATTCCTCGCTGGAGCTCAAGGACAG
GTCGTCGGAGGTAACGTAGTTGGTGAGTTAATGGCGGCGGGGCCGTAATGGTCATGGCA
GCGTCTTTTACAAACGTGGCTTACGAAAGGTTGCCTTTGGACGAGCATGAGGAGCACTTG
CAAAGTGGCGGCGGCGGAGGTGGAGGAATATGTACTCGGAAGCCACTGGCGGTGGCGGA
GGGTTGCCTTTCTTTAATTTGCGGATGAGTATGCCTCAGATTGGAGTTGAAAGTTGGCAG
GGGAATCAGCGCGGCGCGGTAGGGCTCCGTTTTAGCAATTTAAGAAACTTTAATTGTTT
TTTCCACTTTTGTGTTTTCTCCGAATTTATGAAATTATGATTAAAGAAAAAACGAT
ATTGTTTCATGTATTGACCTCTTACTGTCATGGTTTCTTCTATTGGGTTAATTGGCTAGCT
CATAAGAATTGTTTAATTTGGTTATTGTCATCAAATTTGCCACATATAAAGCTTCTAGC
AAAT

>G1075 Amino Acid Sequence (domain in AA coordinates: 78-85)
MAGLDLGTTSRYVHNVDGGGGGQFTTNDNHEDDGGAGGNHHHHHHNHNHQQGLDLIASND
NSGLGGGGGGGSDLVMRPRGRPAGSKNKPPIVITRESANTLRAHILEVSGCDVFE
CISTYARRRQRGICVLSGTGTVNVSIRQPTAAGAVVTLRGTFEILSLSGSFLPPPAPPG
ATSLTIFLAGAQGVVGGNVGELMAAGPVMVMAASFNTVAYERLPLDEHEEHLQSGGGG
GGGNMYSEATGGGGGLPFFNLPMSPQIGVESWQGNHAGAGRAPP*

>G1266 (62..718)
CAATCCACTAACGATCCCTAACCGAAACAGAGTAGTCAAGAAACAGAGTATTTTTCTA
CATGGATCCATTTTAATTAGTCCCATTTCTCCGGCTTCTCACCAGGAATATTCTATCGG
ATCTTCTCCAGATTCTTCTCATCTCTTCTTCTAACAATTACTCTCTTCCCTTCAACGA
GAACGACTCAGAGGAAATGTTTCTTACGGTCTAATCGAGCAGTCCACGCAACAAACCTA
TATTGACTCGGATAGTCAAGACCTTCCGATCAAATCCGTAAGCTCAAGAAAGTCAGAGAA
GTCTTACAGAGGCGTAAGACGACGCCATGGGGGAAATTTCGCGGCGGAGATAAGAGATTCT
GACTAGAAACGGTATTAGGGTTTGGCTCGGGACGTTTCGAAAGCGCGGAAGAGGCGGCTTT
AGCCTACGATCAAGCTGCTTCTCGATGAGAGGGTCTCGGCGATTCTCAATTTTTCGGC
GGAGAGAGTTCAAGAGTCGCTTTCGAGATTAAATATACCTACGAGGATGGTTGTCTCC
GGTTGTGGCGTTGAAGAGGAAACACTCGATGAGACGGAGAATGACCAATAAGAAGACGAA
AGATAGTGACTTTGATCACCCTCCGTGAAGTTAGATAATGTAGTTGTCTTTGAGGATTT
GGGAGAACAGTACCTTGAGGAGCTTTTGGGGTCTTCTGAAATAGTGGGACTTGGTGAAA
GATTAGGATTTGTATTAGGGACCTTAAGTTTGAAGTGTTGATTAAATTTTAACCCTAATA
TGTTTTTTGTTTGTCTTAAATATTGATTCTATTGAGAAACATCGAAAACAGTTTGTATGT
ACTTTTGTGATACTTGGCG

>G1266 Amino Acid Sequence (domain in AA coordinates: 79-147)
MDPFLIQSPFSGFSPEYSIGSSPDSFSSSSNNYSLPFNENDSEEMFLYGLIEQSTQQTY
IDSDSQDLPIKSVSSRKSEKSYRGRVRRRPWGKFAAEIRDSTRNGIRVWLGTFFESAEEAAL

AYDQAAFSMRGSSAILNFSAEVQESLSEIKYTYEDGCSPPVALKRKHSRRRMTNKKTK
DSDFDHRSVKLDNVVVFEDLGEQYLEELLGSSSENSGTW*

>G1311 (41..757)

AAGTATAATAACACAAAGAAACAGAGTAAAGAAAGAAAAATGGATTTTAAGAAGGAAGA
AACAC'TTCGTAGAGGGCCATGGCTCGAAGAAGAAGACGAACGGCTAGTGAAGGTCATTAG
TCTTTTGGGAGAACGTCGTTGGGATTCTTTAGCAATAGTTTCCGTTTGAAGAGGAGTGG
TAAGAGTTGCAGGCTAAGGTGGATGAACCTATCTGAATCCGACTCTGAAGCGTGGACCGAT
GAGTCAAGAAGAAGAGAGAACAATCTTTCAGCTCCATGCTCTATGGGGTAACAAGTGGTC
GAAGATTGCGAGAAGATTACCCGGTAGGACTGATAACGAGATAAAGAACTATTGGAGAAC
TCATTATAGAAAGAAACAGGAAGCTCAAACTATGGAAAGCTCTTTGAGTGGAGAGGAAA
TACAGGAGAAGAATTGTTGCACAAGTATAAGGAAACAGAGATCACTAGGACAAAGACGAC
GTCTCAAGAACATGGTTTGTGTAAGTTGTGAGCATGGAAAGTGGTAAAGAAGCCAACGG
TGGTGTGTTGGTGAAGAGAAAGCTTCGGTGTATGAAATCACCGTATGAAATCGGATTTC
GGATTGGATATCAGAGATTTCTACTGACCAGAGTGAAGCAAATCTTTCAGAAGATCACAG
CAGCAATAGCTGCAGTGAAGACAATATTAACATTGGTACTTGGTGGTTTCAAGAGACTAG
GGACTTTGAGGAGTTTTCATGTTCTCTATGGTCATAATTCTAAAGTTGGTTTATTTACTT
TTTAAAAA

>G1311 Amino Acid Sequence (domain in AA coordinates: 11-112)

MDFKKEETLRRGPWLBEEDERLVKVISLLGERRWDSLAIIVSGLKRSKSCRLRWMNYLNP
TLKRGPMSQEEERIIFQLHALWGNKWSKIARRLPGRTDNEIKNYWRTHYRKQEAQNYGK
LFEWRGNTGEBLLHKYKETEITRTKTTTSQEHGFVEVVSMSGKEANGGVGGRESFGVMKS
PYENRISDWISEISTDQSEANLSESHSSNSCENNINIGTWFPQETRDFFEFSCSLWS*

>G1321 (72..803)

GTTCTTGATTGGTTTGGATCGGTATACTTAGTTGATTACGTAATTAAATAGATCGGCGT
GAAGAAGAAAAATGATCATGTGCAGCCGAGGCCATTGGAGACCAGCTGAAGACGAGAAGC
TCAAGGATCTTGTGCAACAATACGGTCCCTCACAATTGGAACGCCATTGCTCTCAAGCTTC
CTGGTCGCTCTGGTAAGAGTTGTAGATTGAGATGGTTTAATCAATTGGATCCAAGGATCA
ACCGAAACCCCTTTCACGGAAGAAGAAGAAGAAAGACTTTTAGCGGCTCATCGGATCCATG
GGAACAGATGGTCCATCGCATCGCAAGGCTTTTCCCTGGAAGAACTGATAACGCCGTCAAGA
ACCATTGGCACGTCATCATGGCTCGTCGCACACGCCAAACCTCTAAGCCTCGTCTTCTTC
CCTCGACGACTTCGTCCTTCTTTAATGGCGAGTGAACAAATCATGATGAGTTCTGGTG
GTTATAATCATAATTATAGTTCCGATGATCGGAAGAAAAATTTCCAGCAGACTTTATAA
ATTTCCCTTACAAATCTCTCATATCAATCATCTTCACTTCTTAAAGGAGTTTTCCTCCG
GAAAGATCGCTTTAAGTCACAAAGCAAATCAGAGTAAGAAGCCTATGGAGTTCTACAATT
TTCTACAAGTAAACACAGATTCAAAACAAGAGCGAGATTATAGATCAAGATTCAAGTCAAA
GCAAACGCGAGTGAATCGGACACCAACATGAAAGTCATGTTCCATTCTTCGACTTTTAT
CCGTTGGAAGACTCTGCCTCTAGGATTAGTTTTTTTGCAGTAACCTTAAATTTCTAGAT
TAATATTTAGTCCGTATACGTACGAGATTATCTAGGTCGTTAGCATGTATGCTTGATGT
GTATAATCACTAAGTAGTGAGCTATTACCTGCGAAAATTGTAAGAAAAATACATAATGTT
GATGTATCACACATTCTCAATGTCTGTAAATTTCCATCGAGTTGTTAACTATCAAGTT
ATCCGTTTGAAAAA

>G1321 Amino Acid Sequence (domain in AA coordinates: 4-106)

MIMCSRGRHWRPAEDEKLKDLVEQYQPHNWNIAIKLPGRSGKSCRLRWFNQLDPRINRNP
FTEEEERLLAAHRIHGNRWSIIARLFPGRTDNAVKNHWHVIMARRTRQTSKPRLLPSTT
SSSLMASEQIMSSGGYNHNYSSDRKKIFPADFINFPYKFSHINHLHFLKEFFPGKIA
LSHKANQSKKPMFYNFLQVNTDSNKSEIIDQDSGQSKRSDSTKHESHVPPFDLSVGN
SAS*

>G1326 (32..784)

CGACGGTACGGTGGAGATAGAGATAGCATCCATGGAGATGTCTAGAGGAAGCAACAGTTT
TGACAATAAGAAGCCTAGTTGCCAAAGAGGTCACTGGAGACCTGTTGAAGATGACAATCT
CCGGCAACTCGTTGAACAATACGGTCCCAAGAACTGGAATTTTATGCTCAACATCTCTA
TGGAAGATCAGGGAAGCTGTAGATTAAGATGGTACAACCAACTTGATCCAACATCAC
CAAGAAACCCCTTACCAGAGGAGGAAGAAGAGAGACTGCTTAAAGCTCATCGGATCCAAGG
GAATCGTTGGGCCCTCCATAGCCCGACTGTTCCCCGGGAGGACCGACAACGCTGTCAAAA
CCATTTTCATGTATCATGGCTAGACGCAACGGGAAAACCTTCTTCCACAGCTACTTC
TACGTTCAACCAAACTTGGCATACTGTTTGGAGCCCTAGTTCTAGTCTTACAAGGCTAAA
TAGATCCCATTTCCGGCTATGAGGTATCGAAAGGATAAGAGTTGCGGTCTCTGGCCTTA

ATTTTCAGCCTTTTGCCTTCCTTAAACCTCTCATATGGAATC
>G1326 Amino Acid Sequence (domain in AA coordinates: 18-121)
MEMSRGSNSFDNKKPSCQRGHWRPVEDDNLRLQLVEQYQGKNNWFIAQHLYGRSGKSCRLR
WYNQLDPNITKKPFTHEEEERLLKAHRIQGNRWASIARLFPGRTDNAVKNHFHVIMARRK
RENFSSTATSTFNQTWHTVTLSPSSSLTRLNRSHFGLWRYRKDKSCGLWPYFVSPPTNGQ
FGSSSVSNVHHEIYLERRKSKELVDPQNYTFHAATPDHKMTSNEDGSPMGDDGEKNDVTF
IDFLGVGLAS*

TCCTTCCCAAAAATCTTTTAAATTTTATCTGAAAAAATTAAAAACAACCGAAACAAAAAAAT
 AAAATCAAAAATCAAAAATCTCATACCTTCCCTTGCTCTGTATTTTTTCTCTCTCACTAA
 ATCTCCTCATGGATCCTTCTCTCTCTGCAACCAATGATCCTCATCATCCTCCTCCTCTCA
 GTTCACATCTTTCCCTCCTTTACCAACACCAACCCCTTCGCCCTCTCCAACCCACCCCTT
 CTTACCCGGACCCACCGCCGTCGCCGCCGCCAAACAACATCCATCTCATCAAGCAGCTCC
 TCCGCAGCAGCCACAAAACATCTCCAGTTCCTCCTCATCCATCTATTTCCCACCTTCTTA
 CTCTGACATGATTGACCGGGATTCGACGGTTAAACGAACAGGATGGGTCAAGCAAGCA
 AGCTATTTGAGGTACATAGAGAGAATTTACACTGGGATTTCCTACTGCTCATGGAGCTTT
 GTTGACACACCATCTCAAGACTTTGAAGACCAGTGGGATTCTTGTCTATGGTTAAGAAATC
 TTACAAGCTTGCTTCTACTCCTCCTCCTCCTCCTTACTAGTGTAGCTCCTAGTCTTTGA
 ACCTCCAGATCTGATTTTCATAGTCAACGAGAACCAACCTTTACCTGATCCGGTTTGGC
 TTCTTCTACTCCTCAGACTATTAAACGTGGTCTGTGGTCGACCTTCAAAGCTAAACAGA
 TGTGTGTTCAACCTCAACCTCTGACTAATGGAACACTCACCTGGGAACAGAGTGAATTACC
 TGTCTCTCGACAGAGGAGATACAGATACGACCGCCACAGTTACCGTTACAGCCACAGCA
 CGCGGTTAAGAGACCGCCCGGGTCTGTCTAGAAAAGATGGAACCTTCGCCGACGGTGAAGCC
 AGCTGCTTCTGTTTCCGGTGGTGTGGAGACTGTGAAACGAAGAGGTAGACCTCCGAGTGG
 AAGAGCTGCTGGGAGGGAGAGAAAGCCTATAGTAGTCTCAGCTCCAGCTTCAGTGTTTCCC
 GTATGTTGCTAATGGTGGTGTAGACGCCGAGGGAGACCAAGAGAGTGTACGCTGGTGG
 TGCTTCTCTGTTGTCTCACCACACCCACCAACCTTACAGTGTGAGAGTGGAGGAGGGA
 GGTGTGAGTCAAGAAACGAGGAAGAGGACGCGCTCCTAAGATTGGAGGTGTTATCAGGAA
 GCCTATGAAGCCGATGAGAAGCTTTGCTCGTACTGGAACCCCGTAGGAAGACCCAGAAA
 GAATGCGGTGTCTAGTGGGAGCTTCTGGACGACAAGATGGTGACTATGGAGAACTGAAGAA
 GAAGTTTGAGTTGTTTCAAGCGAGAGCTAAGGATATTGTAATTGTGTTGAAATCCGAGAT
 AGGAGGAAGTGAAATCAAGCAGTGGTTCAAGCCATACAGGACCTGGAAGGGATAGCAGA
 GACAAACAAACGAGCCAAAGCACATGGAAGAAGTGCAGTGCACAGCAGGAACACCTTGA
 AATCCGAACAGAGAAGCAGGGTCAAGGACAGCAGAGAAGCAGAGGCAATGCAAGAAGCTCT
 TAACTTAAAGATAAGGACCTTGACATAAAAAGCTAGCAAGTGGTGGGTTTACTTGTGTGTG
 TTACATGAAATTTTAAATCTTATAAGGGTGTTTGCGAGAGAAAAACAAAAGAACAAATGT
 GATGAACTGATGATGATGATTGTGTCTTAACCAAAACAACAGGAGAGGTAGGGTAAATGT
 CTGTAAAGTGAATTAGGATGTTACCATGTGTTTCATGCTTCCCATCTCTCCATCGTCCAT
 ATCTGTGTAGGCACCTTTGTTCTTTGTTCCCTCGTGTTTTTTTAGACTGTTGTGTCTCT
 TATTTCTATTTTGTCTCTTACGCTTTTTAGGAGTTGTTGTTGATGTTTATCAAAAACGCT
 TATGTAATTTTTATGACCACTTCTACTTTTTATGATGGTTTCTT

MDPSLSATNDPHHPPPQQTSFPFFTNTNPFASPNHPFFTGP TAVAPPNNIHLYQAAPPQ
 QQTSPVPHPHSISHPPYSDMICTAIAALNEPDGSSKQAI SRYIERIYTGIPTAHGALLT
 HHLKTLKTSGLVMVKKSYKLASTPPPPPTSVAPSLPPRSD FIVNENQPLPDPVLASS
 TPQTIKRGGRPPKAKPDVVQPQLTNGKLTWEQSEL PVS RPEEIQIQPPQLPLQPQPV
 KRPPGRPRKDGTSPTVKPAASVSGGVETVKRRGRPPSGRAAGRERKPIVVSAPASVFPYV
 ANGGVRRRGRPKRVDAGGASSVAPPPPPPTNVSEGGEEVAVKKRGRGRPKPIGGVIRKPM
 KPMRSFARTGKPVGRPRKNVSVGASGRQDGDYGLKKFEL FQARAKDIVIVLKSEIGG
 SGNQAVVQAIQDLEGIAETTNEPKHMEEVQLPDEEHLETEPEAEGQGQTEAEAMQEALF*
 >G1386 (89..673)

AATTTTATTTCTTCTCTCAAATCTTCCCACCAAAAATTAACCTCTTTCGTTTCACACTAAG
 TCCCTTTTAAAAAGAAAATATCCCAATTAATGGAACGTGACGACTGCCGAGATTTTCAGGA
 CTCGCCGGCGCAGACGACGAGAGAGAAGAGTGAAATATAAAACCAAAGAAGAAAAGAGCCAA
 AGATGATGATGATGAGAAAAGTTGTTTTCGAAGCATCCAAATTTTCGAGGTGTCAGAATGAG
 ACAATGGGAAAAATGGGTGTCCGAAATCAGAGAGCCAAAAAAGAAATCAAGAATCTGGCT
 CGGTACTTTCTCCACGGCGGAGATGGCGGCGGTGCTCACGACGTGGCAGCTTTAGCCAT
 CAAAGGCGGTTCTGCACATCTCAACTTCCCGGAGCTCGCTTATCACCTCCCTAGACCAGC
 TAGTGCCGACCTAAAGACATCCAAGCTGCCGCCGCCGAGCTGCAGCCGCTGTGGCCAT
 TGACATGGATGTAGAGACGTCTTCCGCGTCCGATCTCCACAGTTACGGAACGTCATC
 TCCGCTATGATAGCACTCTCCGACGACGCGTCTCCGATCTTCTGATCTTGTCTCAA
 CGTGAACCATAACATCGATGGCTTCTGGGACTCTTTTCCCTATGAAGAACCCTTCTCTC
 TCAAAGTTACTAGAACTCAAACTATGTCTGTTTTGTATGTATTTTTGTCTGTGACCA
 TTTTTTGACGTCGAAAATCACCCGATAATCCAAATTGTATGATTTATTAATGGTTGATG
 ATTTCTTTTGTGTGGAACAATGTGTATGATACGTAATCAAAAGTTCAAAAAAAAATAAA
 AAAAA

>G1386 Amino Acid Sequence (domain in AA coordinates: TBD)
 MERDDCRRFQDSPAQTERRVKYKPKKKRAKDDDEKVVSXHPNFRGVRMRQWGWVSEI
 REPKKKSRIWLGTFSTAEMAARAHDAALAIKGSAGHLNFPDELAYHLPRPASADPKDIQA
 AAAAAAVIDMDVETSSPSPPTVTETSSPAMIALSDDAFSDLPDLLLNVDNIDGFW
 DSFPYEEFPLSQSY*

>G1421 (292..1155)
 GAAATTTCATCCCTAAATAAGAAAAAGCATCTCCTTCTTTAGTGTCTCTCTTCACCAA
 CTCTTGATTCCATAAGCATATATTAAGAAAGCTCTCTGCTTTCTTCAACTTTCCCGGAA
 AATCTTCTTGTACAAAGCATCAATCTCTTGTGTTTACCAATTTCTCTCTTTATTCCTTT
 TTTGCCCTTTACTTTTCTTAACCTTTGGTCTTTATATATAAACACACGACACAAAGAAGAA
 CACACATAAGTTAAACTATTACAACAGTTTTAAAGAGAGAGATTTAAAAATGGAGACA
 GAGAAGAAAGTTTCTCTCCCAAGATCTTACGAATCTCTGTTACTGATCCTTACGCAACA
 GATTCTGTAAGCGACGAGAAGAAGAGAGTTGATTTGATGCATTATCTACAAAACGACGT
 CGTGTTAAGAAGTACGTGAAGGAAGTGGTGTGATTCCGTGGTTTCTGATAAAGAGAAG
 CCGATGAAGAAGAAGAGAAGAAGCGCGTTGTTACTGTTCCAGTGGTTGTTACGACGGCG
 ACCGAGGAAGTTTCTGTTGAGTGAGGCAAGACCGTGGGGAATGGCGCGCGGAGATTAGA
 GATCCGAGTAGACGTGTTAGGGTTTGGTTAGGTACTTTTGACACGGCGGAGGAAGCTGCC
 ATTGTTTACGATAACGCAGCTATTACGCTACGTGGTCTAACGCAGAGCTTAACCTCCCT
 CCTCTCCGGTGACGGAGAATGTTGAAGAAGCTTCGACGGAGGTGAAGAGGTTTCGGAT
 TTTATCATTTGGCGGTGGAGAATGCTTCTGTTCCGCGGTTTCTGTTCTCGAATCTCCGTT
 TCCGGCGAGTCTACTGCGGTTAAAGAGGAGTTTGTGCGGTGATCGACGGCGGAGATTGTG
 GTTAAAAAGGAGCCGCTTTTAAACGGTTCAGATTTCTCGGCGCCGTTGTTCTCGGACGAC
 GACGTTTTTGGTTTCTCGACGTGATGAGTGAAAGTTTCGGCGGCGATTTATTTGGAGAT
 AATCTTTTGGCGATATGAGTTTGGATCCGGGTTTGGATTCCGGTCTGGGTCTGGATTC
 TCCAGCTGGCACGTTGAGGACATTTTCAAGATATTGGGGATTTATTCGGGTCCGATCCT
 GTCTTAACGTGTTAAGAAATAACTGGCCGTTTAAACGGCGTTTGTGAAAGTTTGTGTTACCG
 GCGACGGCGAGGATTAAAAAAAACGGCGATTTATTTTGAATGAAGATTGTTAAATA

>G1421 Amino Acid Sequence (domain in AA coordinates: 74-151)
 METEKKVSLPRILRISVTDPYATDSSSDEEEVDFDALSTKRRRVKKYVKEVVLDSVSD
 KEKPMKKKKRKRVTVPVVVTATRKFRGRVQRPGWKWABIRDPSSRRVRVWLGTFTAE
 EAAIVYDAAIQLRGPNALNFPPTVTENVVEASTEVKGVSDFIIGGGECLRSPVSVLE
 SPFSGESTAVKEEFVGVSTAEIVVKKEPSFNGSDFSAFLFSDDDVFGFSTSMSESFGGDL
 FGDNLFDMSFGSGFGSGSGFSSWHVEDHFQDIGDLFGSDPVLTV*

>G1453 (39..917)
 CGTCGACGCGAAATAAATCCTAGAAAATAACTATCAATATGATGAAGGTTGATCAAGATT
 ATTCTGTGTAGTATACCGCTGGATTTAGGTTTCATCCGACAGATGAAGAACTTGTCTGGAT
 ATTATCTCAAGAAGAAAATCGCCTCCAGAGGATTGATCTCGACGTTATCAGAGAAATTG
 ATCTTTACAAGATCGAACCATGGGATCTACAAGAGAGATGTAGGATAGGGTACGAGGAGC
 AAACGGAGTGGTATTTCTTCAGCCATAGAGACAAGAAGTATCCGACTGGGACTAGGACAA
 ACCGAGCCACCGTGGCCGTTTCTGGAAGCAACGGCGCGGACAAAGGCGGTTTACCTCA
 ACTCAAACCTTATCGGTATGAGAAAAACGCTTGTCTTTTACCGAGGTCGAGCGCCTAATG
 GCCAAAAGTCCGATTGGATCATTCACGAATACTACAGCCTCGAGTCACACCAGAATCTC

CTCCACAGGAAGAAGGATGGGTAGTGTGTAGAGCATTTAAGAAACGAACGACCATCCCAA
CAAAAAGGAGGCAACTTTGGGATCCGAAC TGCTTATTTCTACGACGACGCACTCTCTTGG
AACCTCTCGACAAGCGAGCCAGACATAATCCTGATTTTACCGCCACACCGTTCAAGCAAG
AACTACTCTCCGAGGCCAGTCACGTCCAGGATGGAGATTTCGGATCTATGTACCTTCAAT
GCATCGATGATGATCAATTCTCCAGCTTCCTCAGCTCGAGAGCCCTCTCTTCCGTCGG
AAATAACTCCCCATAGTACTACTTTTCTGAGAACAGTAGCCGAAAGATGACATGAGCT
CCGAGAAGAGGATCACTGACTGGAGATATCTAGATAAGTTCGTGGCGTCTCAATTTTGA
TGAGTGGAGAAGACTAAAAAAGGCTTTCCTATGCATGCATGCACTAGAAACGTCGTGCGA
TTTTGGATTTACATGCGGCCGCT

>G1453 Amino Acid Sequence (conserved domain in AA coordinates:13-160)
MMKVDQDYSISIPPGFRFHPTDEELVGYYLKKKIASQRIDLVDVIREIDLYKIEPWLQER
CRIGYBEQTEWYFFSHRDKKYPTGTRTNRATVAGFWKATGRDKAVYLNLSKLGMRKTLVF
YRGRAPNGQKSDWIIHEYSSLESHQNSPPQEEGWVVCRAFKKRTTIPTKRRQLWDPNCLF
YDDATLLEPLDKRARHNPDTATPFKQELLSEASHVQDGFSGMYLQCIDDDQFSQLPQL
ESPSPSEITPHSTTFSENSSRKDDMSSEKRITDWRYLDFKVASQFLMSGED*

>G1560 (120..1340)
ATCCTTTCAATTTCCACTCTCTCTAATATAATTACATTTTCCCACTATTGCTGATTCA
TTTTTTTTTGTGAATTATTCAAACCCACATAAAAAAATCTTTGTTTAAATTTAAACCA
TGGATCCTTCATTTAGGTTTCATTAAAGAGGAGTTTCTGCTGGATTCACTGATTCTCCAT
CACCACCATCTTCTTCTTCATACCTTTATTATCTTCCATGGCTGAAGCAGCCATAAATG
ATCCAACAACATTGAGCTATCCACAACCATAGAAAGTCTCCATGAATCAGGGCCACCTC
CATTTTTGACAAAGACATATGACTTGGTGGAAAGATTCAAGAACCAATCATGTGCTGTCTT
GGAGCAAATCCAATAACAGCTTCATTGTCTGGGATCCACAGGCCCTTTCTGTAACTCTCC
TTCCAGATTCTTCAAGCACAATAACTTCTCCAGTTTGTCCGCCAGCTCAACACATATG
GTTTCAGAAAGGTGAATCCGGATCGGTGGGAGTTTGCAAACGAAGGGTTTCTTAGAGGGC
AAAAGCATCTCCTCAAGAACATAAGGAGAAGAAAAACAAGTAATAATAGTAATCAATGC
AACAACTCAAAGTTCTGAACAACAATCTCTAGACAATTTTGCATAGAAGTGGGTAGGT
ACGGTCTAGATGGAGAGATGGACAGCCTAAGGCGAGACAAGCAAGTGTGATGATGGAGC
TAGTGAGACTAAGACAGCAACAACAAAGCACCAAAATGTATCTCACATTGATTGAAGAGA
AGCTCAAGAAGACCGAGTCAAAACAAAACAATGATGAGCTTCTTGCCTGCGCAATGC
AGAATCCAGATTTTATTTCAGCAGCTAGTAGAGCAGAAGGAAAAGAGAAAGAGATCGAAG
AGGCGATCAGCAAGAAGAGACAAAGACCGATCGATCAAGGAAAAGAAATGTGGAAGATT
ATGGTGATGAAAGTGGTTATGGGAATGATGTTGCAGCCTCATCTCAGCATTGATTGGTA
TGAGTCAGGAATATACATATGGAACATGTCTGAATTCGAGATGTGCGAGTTGGACAAAC
TTGCTATGCACATTCAAGGACTTGGAGATAATTCCAGTGCTAGGGAAGAAGTCTTGAATG
TGGAAAAGGAAATGATGAGGAAGAAGTAGAAGATCAACAACAAGGGTACCATAAGGAGA
ACAATGAGATTTATGGTGAAGGTTTTTGGGAAGATTGTTAAATGAAGGTCAAAATTTTG
ATTTTGAAGGAGATCAAGAAAATGTTGATGTGTTAATTGAGCAACTTGGTTATTTGGGTT
CTAGTTACACACTAATTAAGAAGAAATGAAATGATGACTACTTTAAGCATTGTAATCA
ACTTGTTTCTTATTAGTAATTTGGCTTTGTTTCAATCAAGTGAGTCGTGGACTAAGTGC
>G1560 Amino Acid Sequence (domain in AA coordinates: 62-151)
MDPSFRFIKEEFPAGFSDSPSPSSSSSYLYSSSMAEAAINDPTTLYSYPQPLEGLHESGPP
PFLTCKTYDLVEDSRINHVSVWSKSNNSFIVWDPQAFSVTLPRFFKHNNFSSFVRQLNTY
GFRKVNPRDRWEFANEGFLRGQKHLKKNIRRRKTSNNSNQMQPQSSEQQSLDNFCIEVGR
YGLDGEMLSLRRDKQVLMMLVRLRQQQSTKMYLTLIEEKLKKTESKQKQMMMSFLARAM
QNPDFIQQLVBEQKEKRKEIEEAISKRRQRPIDQKRNVEDYGDESGYNDVAASSALIG
MSQEYTYGNMSEFEMSELDKLAHMIQGLGDNSSAREEVLNVEKGNDEBEVEDQQQGYHKE
NNEIYEGEFWEDLLNEGQNFDFEGDQENVDLIQQLGYLGSSSHTN*

>G1594 (1..984)
ATGGATGGAATGTACAATTTCCATTTCGGCCGGTGATTATTACAGATAAGTCGGTTCTGATG
ATGTCACCGGAGAGTCTCATGTTTCTTCCGATTACCAAGCTTTGCTATGTTCTCCGCC
GGTGAAAATCGTGTCTCTGATGTTTTCGGATCCGACGAGCTACTCTCAGTAGCCGTCTCC
GCTTGTGTCGTGCGAGCGGCTTCGATCGCTCCGAGATCCGAAGAAATGATGATAACGTT
TCTCTAACTGTCATCAAAGCTAAAATCGCTTGTATCTCTCGTATCCTCGCTTACTTCAA
GCTTACATCGATTGCCAAAAGGTCCGAGCACCACCGAGATAGCGTGTCTTACTAGAGGAG
ATTCAACGGGAGAGTGATGTTTATAAGCAAGAGGTTGTTCTTCTTCTTGGAGCT
GATCTGAGCTTGATGAATTTATGGAAACGTACTGCGATATATTAGTGAAATACAAATCG

GATCTAGCAAGACCGTTTGACGAGGCAACGTGTTTCTTGAACAAGATTGAGATGCAGCTA
 CGGAACCTATGTACTGGTGTGAGTCTGCCAGGGGAGTTTCTGAGGATGGTGTAATATCA
 TCTGACGAGGAAGTGAAGGTGATCATGAGGTAGCAGAGGATGGGAGACAAAGATGT
 GAAGACCGGGACCTCAAAGATAGGTTGCTACGCAAATTTGGAAGCCGTATTAGTACTTTA
 AAGCTTGAGTTCTCAAAGAAGAAGAAGAAAGGAAAGTTACCAAGAGAAGCAAGACAAGCT
 CTCTTGATTGGTGGAACTCCATTATAAGTGGCCTTACCCTACTGAAGGAGATAAGATA
 GCATTAGCTGATGCAACGGGGTTAGACCAAAAAACAAATCAACAATTGGTTTATAAACCAA
 AGGAAACGTCATTGGAAGCCATCAGAGAATATGCCTTTTCGCTATGATGGATGATTCTAGT
 GGATCATTCTTTACCGAGGAATGA

>G1594 Amino Acid Sequence (conserved domain in AA coordinates: 343-308)

MDGMYNFHSAGDYSKSVLMMSPELSMFPSPDYQALLCSSAGENRVSDVFGSDELLSVAVS
 ALSSEAASIAPEIRNDDNVS LTVIKAKIACHPSYPRLLQAYIDCQKVGAPPEIACLLLEE
 IQRESDVYKQEVVPSSCFGADPELDEFMETYCDILVKYKSDLARPFDEATCFLNKIEMQL
 RNLCTGVESARGVSEDGVISSDEELSGGDHEVAEDGRQRCEDRDLKDRLLRKFGSRISTL
 KLEFSKKKKKKGLPREARQALLDWNHLHYKWPYPTEGDKIALADATGLDQKQINNWFINO
 RKRHWKPSENMPFAMDDSSGSFFTEE*

>G1750 (94..1101)

CCCTTTTCCTCTCTTCTCCAAATCTCTGAAAATTTTACCAGAAATCTCTGTTCTTTTTT
 TCACCAGAATCTCTCTGTTTAAAAATAATAGGTGATGATGATGGATGAGTTTATGGATCTT
 AGACCAGTGAAGTACACAGAGCACAAAGACTGTTATCAGAAAGTACACTAAAAAGTCGTCT
 ATGGAGAGGAAGACCAAGTGTTCGTGACTCGGCCAGGTTGGTTTCGGGTCTCAATGACGGAT
 CGTGACGCCACTGATTCATCAAGCGACGAGGAAGAGTTTCTGTTCCCTCGAAGACGTGTC
 AAGAGATTGATTAACGAGATCAGAGTCGAGCCTAGCAGCTCTTCCACCGGCGACGTCTCT
 GCTTCTCCGACGAAGGACCGGAAAAGAATCAACGTTGATTCTACGGTTCAAAGCCCTCT
 GTTTCGGGCCAAAACAGAGAAGTACCGCGGCGTGAGACAGCGACCATGGGGAAAATGG
 GCGGCGGAGATTCTGTATCCTGAGCAACGCCGAGAAATCTGGCTCGGTACTTTTGCACG
 GCGGAGGAAGCTGCCATCGTCTACGACAACGCAGCAATCAAACCTCGTGCCCTGATGCT
 CTTACCAACTTCACCGTACAACCCAGAACCCAGAACCGGTACAAGAACAAGAACAAGAACCG
 GAGAGCAACATGTCTGGTTTCGATATCAGAATCAATGGACGATTCTCAACATCTATCATCT
 CCGACATCGGTTCTCAACTACCAACATATGTCTCGGAGGAACCAATCGATAGTCTTATC
 AAACCGGTTAAACAAGAGTTTCTTGAACCAGAACAGAGCCAATAAGCTGGCATCTTGGA
 GAAGGTAATACTAATACTAATGATGATTCATTTCCATTGGACATTACATTTCTCGACAAC
 TATTTCAATGAATCATTACCAGACATCTCCATCTTCGATCAACCTATGTCTCCTATTCAA
 CCAACAGAGAAATGATTTCTTCAACGACCTTATGTTATTCGATAGCAACGCAGAAGAATAC
 TACTCTCCGAGATCAAAGAGATTGGTTTCATCGTTCAACGATCTTGATGATTCTTTGATA
 TCCGATCTCTTACTTGTGTGATATTTTTGCCATTAACCAACACCGGTTTGGTTGC

>G1750 Amino Acid Sequence (domain in AA coordinates: 107-173)

MMMDEFMDLRPVKYTEHKTVIRKYTKSSMERKTSVRDSARLVRVSMTRDARDSSDEE
 EFLFPRRRVKRLINEIRVEPSSSSTGTVSASPTKDRKRINVDSTVQKPSVSGQNKYRG
 VRQRPWGKWAAEIRDPEQRRRIWLGTFATAEEAAIVYDNAAIKLRGPDALTNFTVQPEPE
 PVQEQEQEPESNMSVSISESMDDSQHLSSPTSVLNYQTYVSEEPIDSLIKPVKQEFLEPE
 QEPISWHLGEGNTNTNDDSFPLDITFLDNYFNESLPDISIFDQPMSPIQPTENDFFNDLM
 LFDSNAEEYYSSEIKEIGSSFNDLDDSLISDLLLV*

>G1947 (70..918)

ACAACTATTCTCTCCTCTCTCTTTTTTTTATTA AAAAAGCTCAAATTTATATAGGTTTTTT
 GTTCACAAAATGGATTATAACCTTCCAATTCCATTAGAGGGTCTCAAAGAAACGCCACCA
 ACGGCTTTCTTGACGAAAACATACAACATAGTGGAGGATTCAAGCACAAACAACATAGTT
 TCATGGAGCAGAGACAACAACAGCTTCATTGTTTGGGAACAGAGACTTTGCCCTAATT
 TCGCTCCCTAGATGCTTTAAGCACATAATTTCTCCAGCTTTGTTAGACAGCTCAATACT
 TATGGGTTTAAGAAGATTGATACAGAGAGATGGGAATTTGCAAATGAGCATTTTCTGAAG
 GGAGAGAGGCATCTTCTTAAGAACATCAAGAGAAGAAAGACATCATCTCAAACGCAAACG
 CAGTCGCTAGAAGGAGAGATCCATGAGCTGCGAAGAGACAGAATGGCTTTAGAAGTAGAA
 CTGGTTAGACTGCGACGAAAACAAGAAAGCGTGAAGACATATCTGCATTTGATGGAAGAG
 AAAGTGAAGTCAAGAAAGTAAAGCAAGAAATGATGATGAATTTCTTGCTAAAGAAGATT
 AAGAAACCGAGTTTTTTACAGAGCTTAAGGAAACGTAATCTGCAAGGAATCAAGAATCGA
 GAGCAAAAGCAAGAGGTGATCTCAAGCCATGGTGTGAGGATAATGGAAAGTTTGTATAA
 GCTGAGCCAGAAGAGTATGGTGTGATGACATCGATGATCAATGTGGAGGTGTGTTTGATTAT

GGTGATGAGCTTCACATAGCTTCAATGGAGCATCAAGGACAAGGGGAGGATGAAATTGAA
ATGGATAGTGAAGGAATTTGGAAGGGTTTCGTGTTGAGTGAGGAGGAGATGTGTGATTGA
GTGGAACATTTTATATAATAAACTAATGTATTATGAGAGGTTTTTTTTTTGTTTTTTGCT
TTTTTTTTTCCGAGTTTGTCAATCAAGCATGTATACAATTGGGGCCAACTAAAAGCCCAA
CAAAATATTGGCCCTTGGCATTGTGTTAACAAATTGACTAATTCGGCCACACCTTCC

>G1947 Amino Acid Sequence (domain in AA coordinates: 37-120)
MDYNLPIPLEGLKETPPTAFLTKTYNIVEDSSTNNIVSWSRDNNSFIVWEPETFALICLP
RCFKHNNFSSFVRQLNTYGFKKIDTERWEFANEHFLKGERHLLKNIKRRKTSSQTQTQSL
EGEIHLELRDRMALEVELVRLRRKQESVKTYLHLMEEKLKVTEVKQEMMMNLLKKIKKP
SFLQSLRKRLNQGKIKNREQKQEVISSHGVEDNGKFVKAPEEYGGDDIDDQCGGVFDYGDE
LHIASMEHQGQGEDEIEMDSEGIWKGFLVEEEMCDLVEHFI*

>G2011 (309..1547)

AATGTCGGTTGTACAATTATTTGTCACTAAAGTTTCCAAATTTCTTCTAAACTGATGAAT
CAATGGAACATGATGACGAAAAAGATAAATCCACGGTGGCGGGAACGACCCACCCATT
CCACCGCTCTCTATTTCCCGAGATTTTTCATTATCTGACTACAGTTTGTCTCGTTACT
TCCTTCCCTAAACCTTTTATAAACCATTAACCTCTCATCTTCTTCTTAAACCCCTA
ATTATCACACACACCCCAATTTCTCACTCTCTCTCTCACTAAAACCCGTAAATTTTCTAC
TATATCAAATGAGCCCAAAAAAGATGCTGTTTCTAAACCAACTCCAATTTCACTACCCG
TTTCGAGACGATCCGATATACCCGGGTCTCTCTACGTCGACACTGACATGGGTTTCTCTG
GGTCACCACTTTCCCATGCCACTAGACATCTTACAAGGGAATCCAATTTCCACCTTTTTTAT
CCAAGACTTTTGATTGGTTGATGACCCGACTCTTGACCCGGTCATCTCTTGGGGACTGA
CCGGAGCTAGCTTCGTAGTTTGGGATCCTTAGAGTTTGCCAGAATCATACTTCCAAGGA
ATTTCAAACACAACAATTTCTCCAGCTTCGTCAGACAGCTTAACACTTATGGATTTTGGAA
AGATTGATACTGACAAGTGGGAATTCGCTAACGAGGCTTTCTTAGAGGCAAGAAGCATC
TTCTGAAGAACATTCATCGTCGTCGATCACCACAATCCAACCAAACTTGCTGCAGTAGCA
CTAGCCAAAGCCAAGGGTCACCTACTGAGGTTGGAGGAGAGATTGAGAAGCTGAGGAAAG
AGCGGCGTGCATTGATGGAGGAAATGGTTGAGCTTCAGCAGCAAAGCAGAGGCACAGCTC
GACATGTGGACACTGTAAACCAGAGGCTGAAAGCTGCAGAGCAACGTGAGAAGCAATTGC
TCTCTTCTTGGCTAAGTTGTTTTCAGAACCGGGGTTTCTTGAACGCCTGAAGAACTTCA
AAGGAAAAGAAAAGAGGAGCTCTTGGATTGGAAAAGGCGAGAAAAGATTTCATCAAGC
ACCACCAGCAGCCTCAAGATTCTCCAACAGGAGGGGAGGTGGTGAAGTATGAAGCTGATG
ATTGGGAGAGATTGCTAATGTATGACGAAGAGACTGAGAACACCAAGGGTTTAGGAGGGA
TGACTTCAAGCGATCCAAAAGGCAAGAATTTGATGTATCCATCAGAAGAAGAGATGAGCA
AACCAGATTACTTGATGTCTTCCCATCTCTGAAGGACTTATTAACAAGAAGAGACGA
CATGGAGCATGGGTTTCGATACTACAATACCGAGTTTCAGCAACACCGATGCATGGGGAA
ACACAATGGACTATAATGATGTCTCAGAGTTTGGTTTGTCTGCAGAAACAACAAGTGATG
GTTTGCCTGATGTCTGTGCGGAACAATTTGCTGCAAGGAATCAGAGAGACTGGATTCAACT
GGCCAACGTGGTATGATGATGATAATACGCCAATGAATGATCCTTAGGATCTTTTCATAT
ATAGTTTAGACCAAAAACCCGTTTCTTATCGGGTGAACATTAATTCAATTATTCATTTTG
AATGCACTCTTTATACATATATATAATATTGATGAGTTTGATTGTTCCAAAAA

>G2011 Amino Acid Sequence (domain in AA coordinates: 56-147)
MSPKKDAVSKPTPISVPSRRSDIPGSLYVDITDMGFSGLPLMPLDILQGNPIPPFLSKT
FDLVDDPTLDPVISWGLTGASFVWVWDPLEFARIILPRNFKHNNFSSFVRQLNTYGFRRKID
TDKWEFANEAFRLRGKHLKNIHRRRSPQSNQTCCSSTSQSQGSPTGVGGEIEKLRKERR
ALMEEMVELQQQSRGTARHVDTVNQRLKAAEQRQKQLLSFLAKLFQNRGFLERLKNFKGK
EKGGALGLEKARKKFIKHQQPQDSPTGGEVVKYEADDWERLLMYDEETENTKGLGGMTS
SDPKGKNLMYPSEEEEMSKPDYLMFSPSEGLIKQEETTWSMGFDTTIPSPSNTDAWGNTM
DYNDVSEFGFAAETTS DGLPDVCWEQFAAGITETGFNWPTGDDDDNTPMNDP*

>G2094 (1..450)

ATGCTAGATCCCAAGGAGAAAGTAATCGATTGAGAAATCAATGGAAGCAAACCTCACATCA
GTAGATGCGATCGAAGAACACAGCAGTAGCAGTAATGAAGCTATCAGCAACGAGAAG
AAGAGTTGTGCCATTTGTGGTACCAGCAAAACCCCTCTTTGGCGAGGCGGTCCTGCCGGT
CCCAAGTCGCTTTGTAACGCATGCGGGATCAGAAACAGAAAGAAAGAAGAACTGATC
TCAAATAGATCAGAAGATAAGAAGAAGAAGATCATAACAGAAACCCGAAGTTTGGTGAC
TCGTTGAAGCAGCGATTAAATGGAATTGGGGAGAGAAGTGATGATGACGCGATCAACGGCT
GAGAATCAACGGCGGAATAAGCTTGGCGAAGAAGAGCAAGCCCGCTGTACTCATGGCT
CTCTCTTATGCTTCTTCCGTTTATGCTTAA

>G2094 Amino Acid Sequence (domain in AA coordinates:43-68)
MLDPTEKVIDSESMESKLTSDAIEEHSSSSSNEAISNEKKSCAICGTSKTPLRGGPAG
PKSLCNACGIRNRKKRRTLISNRSEDKKKKSHNRNPKFGDSLKQRLMELGREVMMQRSTA
ENQRRNKLGEEEQAAVLLMALSYASSVYA*

>G2113 (90..590)
ATAACAAACTCATCAAACTTCCTCAGCGTTTCTTTTCTTACATAAAACAATTTTCTTAC
ATAACAAATCTTGTGTTTGTGTTGTCATGGCACCACAGTTAAACCGCGGCCGTCA
AAACCAACGAAGGTAAACGGAGTCCGTTACAGAGGAGTGAGGAAGAGACCATGGGGACGTT
ACGCAGCCGAGATCAGAGATCCTTTCAAGAAGTCACGTGTCTGGCTCGGTACTTTCGACA
CTCCTGAAGAAGCCGCTCGTGCCTACGACAAACGTGCTATTGAGTTTCGTGGAGCTAAAG
CCAAAACCAACTTCCCTTGTTCACATCAACGCCCACTGCTTGAGTTTGACACAGAGCC
TGAGCCAGAGCAGCACCGTGGAATCATCGTTTCTTAATCTCAACCTCGGATCTGACTCTG
TTAGTTTCGAGATTCCCTTTTCTAAGATTACGGTTAAGGCTGGGATGATGGTGTTCGATG
AAAGGAGTGAATCGGATTCTTCGTGGTGATGGATGTCGTTAGATATGAAGGACGAC
GTGTGGTTTGGACTTGGATCTTAATTTCCCTCCTCCACCTGAGAACTGATTAAAGATTTA
ATTATGATTATTAGATATAATTAATGTTTCTGAATTGAG

>G2113 Amino Acid Sequence (domain in AA coordinates: TBD)
MAPTVKTA AVKTNEGNVRYRGRVGRYPWGRYAAEIRDPFKSRVWLGTFTDTPEEAARAYD
KRAIEFRGAKAKTNFPCYNINAHCLSLTQSLSQSSTVSESFNPNLNLGSDSVSSRFPFKI
QVKAGMMVFDERSSESSSVMDVVRVYEGRRVLDLNLNFPFPPEN*

>G2115 (41..733)
AATCACTCTACAAAGCCTGTACGTACACAACAACATTACCATGGTGAAACAAGAACGCAA
GATCCAAACCAGCAGCACAAAAAGGAAATGCCTTTGTTCATCATCACCATCTTCTTCTTC
TTCTTCATCTTCTTCTCGTCTTCGTCTTCGTGTAAGAACAAGAACAAGAAGAGTAAGAT
TAAGAAGTACAAAGGAGTGAGGATGAGAAGTTGGGGATCATGGGTCTCTGAGATTAGGGC
ACCAATCAAAAGACAAGGATTGGTTAGGTTCTTACTCAACAGCTGAAGCAGCTGCTAG
AGCTTACGATGTTGCACTCTTATGTCTCAAAGGCCCTCAAGCCAATCTCAACTTCCCTAC
TTCTTCTTCTCTCATCATCTTCTTGATAATCTCTTAGATGAAAATACCCTTTTGTCCCC
CAATCCATCCAAAGAGTAGCTGCTCAAGCTGCCAACTCATTAAACATTTTGCCCCCTAC
TTCATCAGCCGTCTCGTCAACGCTCCGATCATGATCATCACCATGATGATGGGATGCAATC
TTTGATGGGATCTTTTGTGGACAATCATGTGTCTTTGATGGATTCAACATCTTCATGGTA
TGATGATCATAATGGGATGTTCTTGTGTTGATAATGGAGCTCCATTCAATTACTCTCCTCA
ACTAAACTCGACGACGATGCTCGATGAATACTTCTACGAAGATGCTGACATTCGCTTTG
GAGTTTCAATTAATCCGACGGTCCATAATACATACTTTAATTAGT

>G2115 Amino Acid Sequence (conserved domain in AA coordinates:46-115)
MVKQERKIQTSSSTKEMPLSSSPSSSSSSSSSSSSCKNKNKSKIKYKGVMRMRWSGS
WVSEIRAPNQKTRIWLSYSTAEAAARAYDVALLCLKGPQANLNFPTSSSSHLLDNLDD
ENTLLSPKSIQRVAAQAANSFNHFAPTSSAVSSPSDHDHDDGMQSLMGSFVDNHVSLM
DSTSSWYDDHNGMFLFDNGAPFNYSPLNSTTMLDEYFYEDADIPLWSFN*

>G2130 (41..988)
CCTCTCTTCAATTTTAACTCCCTCTCTCTCTCTCTATGGAGAGACGAACGAGACG
AGTGAAGTTCACAGAGAATCGTACGGTCACAAACGTAGCAGCTACACCATCTAACGGGTC
TCCGAGACTGGTCCGTATCACTGTTACTGATCCTTTTCGCTACTGACTCGTCTAGCGACGA
CGACGACAACAACAGTCACGGTGGTTCCAAGAGTGAAACGATACGTGAAGGAGATTAG
ATTCTGCCAAGGTGAATCTTCTTCTCCACCGCGGAGGAAAGGTAAAGCACAAGGAGGA
GGAAAGCGTAGTGGTTGAAGATGACGTGTCGACGTGCGTGAAGCCTAAAAAGTACAGAGG
CGTGAGACAGAGACCTTGGGGAAAATTCGCGGCGGAGATTAGAGATCCGTCGAGCCGTAC
TCGGATTTGGCTTGGGACTTTTGTACGGCGGAGGAAGCTGCTATAGCGTACGATAGAGC
CGCGATTATCTCAAAGGACCTAAAGCGCTACGAATTTCTTAACCTCGCCGACGCCAAC
GCCGGTTATCGATCTCAAACGGGTTTCCGCTGCGATTACGGTAGAGATTCTCGGCAGAG
CCTTCATTACCCGACCTCTGTTCTAAGATTCAACGTCAACGAGGAAACAGAGCATGAGAT
TGAAGCGATCGAGCTATCTCCGAGAGAAAGTCGACGGTTATAAAGAAGAAGAAGAATC
GTGGCGGGTTTGGTGTTCGGGATCCGTATCTGTTACCGGATTATCTCTCGCCGCGCA
ATGTTTTTGGGATACCGAAATGCCCCCTGACCTTTTGTCTTCGATGAAGAAACCAAAAT
CCAATCAACGTTGTTACCAACACAGAGGTTTCGAAACAAGGAGAAACGAAACTGAAGA
TTTCGAGTTTGGTTGATTGATGATTTTCGAGTCTTCTCCATGGGATGTGGATCATTTCTT
CGACCATCATCACTCTTTCGATTAAAAATCTCTTCTTTTGGGGAAATTTTGTG

>G2130 Amino Acid Sequence (domain in AA coordinates 93-160)
MERRTRRVKFTENRTVTNVAATPSNGSPRLVRITVTDPFATDSSSDDDNNVTVVPRVK
RYVKEIRFCQGESSSSTAARKGKHKEESVVVEDDVSTSVKPKKYRGVRQRPWGKFAAEI
RDPSSRTRIWLGTFTVTAEEAAIAYDRAAIHLKGPALTNFLTPPTPVIDLQTVSACDY
GRDSRQSLHSPTSVLRFNVNEETEHEIEAIELSPERKSTVIKEEEESSAGLVFPDPYLLP
DLSLAGECFWDTEIAPDLLFLDEETKIQSTLLPNTTEVSKQGENETEDFEFGLIDDFESSP
WDVDHFFDHHHHSFD*

>G2147 (162..1262)
CTGTGATTGTCAAGAGTTTGAACACACAAAGAAGAAAGAAGAACTCAACATTTCAAGCAA
GAAGAAAGAGAGAAGAGAGAAGGTCCATAATAGAGAGAACAAAAAAGAGAGCTTAA
TTGTCAAGTTTATTCTCTGCAACGTGCGGCCTAAGTAACACATGTCTGAATTATGGAGTTA
AAGAGCTCACATGGGAAATGGGCAACTAACCGTTCATGGTCTAGGCGACGAAGTAGAAC
CAACCACCTCGAATAACCTATTGGACTCAAAGTCTCAACGGTGTGTGAGACTTTGGAGT
CTGTGGTTCATCAAGCGGCTCTACAGCAGCCAAGCAAGTTTTCAGCTGCAGAGTCCGAATG
GTCCAAACCAACAATTATGAGAGCAAGGATGGATCTTGTTCAGAAAACGCGGTTATCCTC
AAGAAATGGACCGATGGTTCGCTGTTCAGAGGAGAGCCATAGAGTTGGCCACAGCGTCA
CTGCAAGTGCAGAGTGGTACCAATATGTCTTGGGCGTCTTTTGAATCCGGTTCGGAGCTTGA
AGACAGCTAGAACCAGGAGACAGAGACTATTCCGCTCTGGATCGGAACTCAAGATACTG
AAGGAGATGAACAAGAGACAAGAGGAGAAGCAGGTAGATCTAATGGACGACGGGACGAG
CAGCAGCGATTACAACGAGTCCGAAAGGAGACGGCGTGATAGGATAAACAGAGGATGA
GAACACTTTCAGAAGCTGCTTCTACTGCAAGTAAGGCGGATAAAGTCTCAATCTTGGATG
ATGTTATCGAACACTTGAAACAGCTACAAGCACAAGTACAGTTCATGAGCCTAAGAGCCA
ACTTGCCACAACAATGATGATTCCGCAACTACCTCCACCACAGTCAGTTCTCAGCATCC
AACACCAACAACAACAACAACAACAGCAGCAGCAGCAGCAACAACAGCAGCAACAGTTTC
AGATGTCTGTTGCTTGCAACAATGGCAAGAATGGGAATGGGAGTGGTGGAAATGGTTATG
GAGGTTTAGTTCCTCCTCCTCCTCCTCCACCAATGATGGTCCCTCCTATGGGTAACAGAG
ACTGCACCAACGGTCTTTCAGCCACATTATCTGATCCATACAGCGCTTTTTCGCACAGA
CAATGAATATGGATCTCTACAATAAATGGCAGCAGCTATCTATAGACAACAGTCTGATC
AAACAACAAGGTAATATCGGCATGCCTTCAAGTCTTTCGAATCATGAGAAAAGAGATT
AGTCTAGCGACCTAGTATTATTGATCCATATATATAGTTCTTGAAAGATTGTTGTATCAT
GATTGTAAAACTGTTTGTAGTATGGAAAAAGACTTGCAGATAAAA

>G2147 Amino Acid Sequence (domain in AA coordinates:160-234)
MSNYGVKELTWENGQLTVHGLGDEVEPTTSNNPIWTQSLNGCETLESVVHQAALQQPSKF
QLQSPNGPNHNYESKDGSCSRKRGYPQEMDRWFVAVQEESHVRVGHVSHTASASGTNMSWASF
ESGRSLKTARTGDRDYFRSGSETQDTEGDEQETRGEAGRSNGRRGRAAAIHNESERRRRD
RINQRMRTLQKLLPTASKADKVSILDDVIEHLKQLQAQVQFMSLRANLPQQMMIPQLPPP
QSVLSIQHQQQQQQQQQQQQQQQQQFQMSLLATMARMGMGGGGNGYGLVPPPPPPPMV
PPMGNRDCNTNGSSATLSDPYSAFFAQTMMNDLYNKMAAAIYRQSDQTTKVNIGMPSSSS
NHEKRD*

>G2156 (384..1292)
TTTTTTTTCCCTTTCCTCGTTCAAAAAAGTACTTGCAGAGTCACTCACTCTCAGTCTCA
GCACATGAATTAATTTGAAGCTTCCCTAGAATTCTTTACATCAATTAATACGACACCGT
CTCGGGTGAAGAATCTCTCCTCTCTTGCCCTAAAGCGAGTTAGGGTTTAAACACACAAAGC
ATACCCTTTAGATTTGTGTCTCTTAGCTCTGTTTTTGTGCGCTTGTGTAAACCGATCAACT
CAAGCTATTGGTCTCCTCACCTCCTGAAATTTGACTTCTCCAATGGATCTCAAAGTTTCTC
TTATATGAATCTATCTTACCCTCACAATATCTTTATATATATGAGCCACAAGAACAAG
AAGAGTCAGTAGATGCGGCTGCCATGGACGGTGGTTACGATCAATCCGGAGGAGCTTCTA
GATACTTTCAACACTCTTCAGGCTTGAGCTTCATCACCAGCTTCAACCTCAGCCTCAAC
TTCACCTTTGCTCAGCCTCAGCCTCAACCTCAGCCTCAGCAGCAGAATTCAGATGATG
AATCTGACTCCAACAAGGATCCGGGTTCCGACCCAGTTACCTCTGGTTCAACCGGGAAC
GTCCACGTGGACGTCTCCGGGATCCAAGAACAAGCCGAAGCCACCGGTGATAGTACTA
GAGATAGCCCCAACGTGCTTAGATCTCATGTTCTTGAAGTCTCATCTGGAGCCGACATAG
TCGAGAGCGTTACCCTTACGCTCGCAGGAGAGGAAGAGGAGTCTCCATTCTCAGTGGTA
ACGGCACGGTGGCTAACGTGAGTCTCCGGCAGCCGCAACAGACAGCGGCTCATGGGGCAA
ATGTTGGAACCGGAGGTGTTGTGGCTCTACATGGAAGGTTTGTAGATACTTTCCCTCAGAG
GTACGGTGTGTCGGCCCTTGCGCCGCCAGGATCCGGTGGTCTTTCTATCTTTCTTCCG
GCGTTCAAGGTGAGGTGATTGGAGGAAACGTGGTGGCTCCGCTTGTGGCTTCGGGTCCAG

>G2156 Amino Acid Sequence (domain in AA coordinates:66-86)
MDGGYDQSGGASRYFHNLFRLPELHHQLQPQPQLHPLPQPQPQPQQNSDDSDSNKDP
GSDPVTSGSTGKRPRGRPPGSKNKPVPVIVTRDSPNVLRSHVLEVSSGADIVESVTTYA
RRRGRGVSIILSGNGTVANVSLRQPATTAHANGGGTGGVVALHGRFEILSLTGTVLPPPA
PPGSGGLSIFLSGVQGVIGGNVAPLVASGPFVILMAASFNTFERLPLEDEGEGGEGG
GEVGEVGGGEGGPPPATSSPPSGAGQGQLRGNMSGYDQFAGDPHLLGWGAAAAAAPPRP
AF*

TCCTCCCTTAATTAGTATCAAAAATGGTGAAAACACTTCAAAGACACCAAAGAGAATGT
CATCTCCATCATCATCATCTTCATCATCCTCATCAACATCATCATCATCCATAAGGATGA
AGAAGTACAAAGGGAGTGAGAATGAGAAGTTGGGGTTCATGGGTTCAGAGATCAGAGCT
CTAATCAAAAGACAAGGATCTGGCTTGGTTCTTACTCAACTGCTGAAGCCGCGGCTAGAG
CCTACGACGCGCAGCATCTCTATGTCTTAAAGGATCCTCAGCTAATAATCTCAACTTCCAG
AGATCTCAACTCTCTCTCTACCATTATCAACAATGGTGATAACAACAATGACATGTCCC
CTAAGTCTATACAAAGAGTAGCAGCTGCAGCTGCTGCTGCCAACACAGATCCTTCTCAT
CATCAGTCTCTACTTCATCTCCATTGCTTTCCTCTCCATCTGAAGATCTCTATGATGTTG
TCTCCATGTACAGATATGACCAACAAGTCTCCTTGCTGAATCATCATATGTTACAAC
GCTTTGATGGTGATGATTCAGTTTCATGTTCAATGAAGATCTCCGCGCGTATGTTGACAA
CATCACTTTCTGATGATTTCTTTGAGGAAGGATATCAGATTATGGAACCTTCTGCTGAT
TCTACTTTTCATTATACCTTATTCTTTG

ATACCAAACTTTTAATGTCAACCACAGAGAATGAAGCTATCATCAACCACCAGTTACCAAC
AACGAACCAACCGCCACCCTTCTGCCGTTAAATCTTGCGGCGGAGGAGGTAAAGAAACC
AGCTCATCGACCAAGGAGCATCCAGTGTACCACGGAGTTTCGCAACGCCGATGGGGAAAA
TGGGTTTCTGAGATCAGAGAGCCCCGAAAAAAGTCTCGGATTTGGCTCGGATCTTTCCG
GTGCCGGAGATGGCTGCTAAGGCCATACGACGTGGCAGCGTTTGTCTAAAAGGTAGAAAA
GCTCAGCTGAATTTCCCTGAAGAACTACGAGGATCTACTTCGACCGTCCACGTGTATCCC
AGAGATGATCCAAAGTCGACGCGGCCAAAGACGCCAACCGGTGAAGATCATCAAAATGGGA
GATGATGACGTGCGAGGAATAGACGACGGAGATGATTTCTGGGAAGGCATTGAGCTGCCT
GAGCTTATGATGAGTGGAGGTGGGTGGTGC CGGAGCCTTTGTGTCCGGAGATGATGCC
ACGTGGCTTGTGACGGAGACTTGTATCAGTATCAGTTATGGCGTGTCTGTGAGTGTG
CTGTCCATTGTGTCGTATTCTGTATACGTGTACGTTGTATCGTTATTGTGTTGGCTCACT
TAATTTAATGCATATGACCTGTATATTTTCATTTATTTGTTTCTAGTTTATTGTTTACGC
GATTAATAATAGATACCTGTTTCTCAAGTTAGTTATCAGGTTTGTATGCATCTACAAAA
ATACGTATAAGTGATGTTCTTATATACAGTTTTTGTGTTGATAAGTATTGCTACTTATT
CTAAAAA

AAATCATAAAAGCCTCTCTCTTAGTCTATTTTATCTCACGGCTCTCTCTCCCTCTCTA
CACACACAAAACCAATAAAGCGTAAAACTGAAATATTTTAATTACAATTAGAAAGAGAA
CATATTAAATATGTCAAATATAACAAAGAAGAAGTGTAATGGAAATGAAGAGGGTGACAG
CAGAGGAAGGGCCTTGGACACTCGAGGAAGACACTCTTCTACCAATTACATTCCCAT

AACGGTGAAGGCCGATGGAATCTGCTCGCTAAATCTTCTGGGCTAAAGAGAGCAGGAAAA
 AGTTGTAGATTGAGATGGTTGAATTACCTTAAACCCGACATAAAGCGTGGGAATCTCACT
 CCTCAAGAACAACCTTTTAACTCCTTGAGCTCCATTCTAAATGGGGTAATAGGTGGTCAAAA
 ATTTTCGAAGTATTTACCAGGAAGAACAGACAACGATATCAAAAACTACTGGAGAACTAGA
 GTCCAGAAACAAGCACGCCAGCTCAACATAGATTCCAATAGCCACAAGTTCATAGAAGTT
 GTTCGTAGCTTTTGGTTTCCAAGACTGATCAACGAGATTAAAGACAACCTCATAACCAAC
 AATATTAAGCTAATGCTCCTGATTTACTTGGACCAATTTTACGAGACAGCAAAGATTTG
 GGTTCACAACATGGATTGTCCACTTCCATGTGAGAAGATCTCAAGAAAACCTTCACAA
 TTCATGGATTTTCTGATCTTGAAACCACAATGTCTTGGAGGATCACGAGGGGGTAGT
 AGTCAATGTGTGAGTGAGTTTATAGCTCCTTCCCTTGCCTAGAGGAGGAGTACATGGTG
 GCCGTTATGGGCAGTTTACAGCATTTTACGATTGTCATGATTGTACGTGGCTGATTCCAAG
 TACGAGGATGATGTGACACAAGATCTAATGTGGAACATGGATGACATTTGGCAGTTTAAAC
 GAGTATGCACACTTTAATTAGGTTATATTATTTATGTACTTCTTACAACCTGGAGGGG
 TTTATCGGTCTTTTATTAAATTTGATTGTTTTGGATTCTTAAAAATGTGTTCTTATTA
 TAGTTTTTAAAGAAAAAATGTTTAAAGCGCAAAAAAATAAAAAAAAAAAAAAAAAAAAA
 >G2893 Amino Acid Sequence (conserved domain in AA coordinates: 19-120)

MSNITKKKCNNGNEEGAEQRKGPWLTLEEDTLTNYISHNGEGRWNLLAKSSGLKRAGKSCR
 LRWLNLYLKPDIKRNLTLPQEQLLILELHLSKWNRSKISKYLPGRITDNDIKNYWRTRVQK
 QARQLNIDSNSHKFIEVVRSFWRFLINEIKDNSYTNNIKANAPDLLGPILRDSKDLGFN
 NMDCSTSMSEDLKKTQFMDFDLETTMSLEGRSGSSQCVSEVYSSFPCLLEEYMVAVM
 GSSDISALHDCHVADSKYEDDVTQDLMWNMDDIWQFNEYAHFN*

>G340 (97..834)

ATGAAATCTCTAGTTTTTTTTTTGTTCCTTTCTTAAATTTTCAAAGAAAGACATTTATT
 AAACCAAAATAACTCTTTAGATCATTGCAAGGAAAAATGTTGAAAAGTGCAAGTCCAATG
 GCATTCTACGATATCGGAGAGCAGCAATACTCTACTTTCCGGGTACATTTTAAGCAAACCT
 GGGAAACGACAGGAGCTTACGAGATTGACCTTTCGATCCCAACATCGACGATGCGATCTAC
 GGCTCAGATGAGTTCCGTATGTACGCTTACAAAATCAAACGGTGTCTCTGACTCGTAGC
 CACGACTGGACGGAGTGTCCCTACGCTCACCCTGGCGAGAAAGCCACACGCCGTGATCCT
 CGCCGTTACACTTACTGTGAGTGCATGCCCGGCTTTCCGAAATGGCGCATGCCACCGT
 GGCGACTCATGCGAATTCGCACATGGCGTATTCGAGTACTGGCTCCACCCGGCGCGTTAC
 CGAACACGCGCATGTAACGCCGGGAACCTGTGTGAGAGGAAAGTGTGTTTCTTTGCCAC
 GCGCCGGAGCAGCTAAGGCAGTCTGAAGGAAAGCACAGGTGCAGGTACGCATATAGCCG
 GTGAGGGCTAGAGGTGGTGGAAACGGCGATGGAGTGACGATGAGAATGGACGACGAGGGT
 TACGACACGTACCGTCTCCGGTGAGAAGCGGGAAAGATGATTTAGATAGTAACGAGGAG
 AAGGTGTTGTTGAAGTGTGGAGTCGGATGAGCATTGTGGATGATCATTATGAGCCGTCC
 GATTTGGATTTGGATTGTGACACTTTGATTGGATCTCAGAGTTGGTTCGATTAAATTTGG
 GAAATCAAAGCAGAGAACAAGAAACCCGATAAATAAAGTGGATTTTGTAAATCCAC
 AAGATCAAGATTCAGATGAGAGATCTTGTGATGTATATGGTAAATTTAATTGTAATGAT
 TTATTGCAATGTCGCAAAAGAAAGTTACTTCTCTTTCATGTAAACAGATTCTTGATCTTC
 TATAAGTCTTTGTATTAA

>G340 Amino Acid Sequence (domain in AA coordinates: 37-154)
 MLKSASPMFYDIGEQYSTFGYILSKPGNAGAYEIDPSIPNIDDAIYGSDEFMYAYKI
 KRCPRTRSHDWTECPYAHREKATRRDPRRYTYCAVACPFRNGACHRGDSCEFAHGVFE
 YWLHPARYRTRACNAGNLCQRKVCFFAHAPEQLRQSEGKHRCRYAYRPVRARGGGNGDGV
 TMRMDDEGYDTSRSPVRSGKDDLSNEEKVLLKCSRMSIVDDHYEPSDLDDLSDLHFDWI
 SELVD*

>G39 (75..638)

GTTTCCACAGTCCCGTACTTGTGCATAAACTGTAAAACTACTCTGAAAATTTTGCT
 TCTGTTAGGATATAATGCCACCCTCTCCTCTAAATCTCCTTTTATTAGCTCTTCACTCA
 AAGGAGCTCATGAAGATCGCAAATTTAAATGCTATAGGGGTGTCCGAAAGAGGTCTTGGG
 GCAAATGGGTGTCTGAAATCAGAGTTCCAAAGACTGGACGACGAATATGGCTAGGTTTAT
 ACGATGCTCCAGAGAAGGCAGCTAGAGCCTATGATGCTGCTTTGTTCTGTATTAGGGGTG
 AGAAGGGAGTTTACAATTTTCCACTGATAAAAAGCCGAGCTTCCAGAAGGTTCTGTCC
 GGCTCTGTCCAAGCTCGACATACAGACAATAGCAACAACTATGCTTCATCAGTTGTGC
 ATGTACCTTCCCATGCCACCACACTCCCGGCAACAACCCAGGTTCCCTCTGAAGTTCCTG
 CTTCCTCTGATGTTTCTGCTTCTACTGAGATTACAGAGATGGTTCGATGAATATTATCTCC
 CAACCGATGCAACTGCAGAATCAATATTCTCAGTTGAAGACTTACAACCTGGACAGTTTCC

TCATGATGGACATTGATTGGATAAAACAATCTAATCTGATGTGTAACGTCACCTTGCAGTGA
CATTTAATATGGTTTANCTATCAGTTACCTGTCTGCTTCTTGTAAAGGTATACTTGGATC
CTTGCTTTTGAACCTGTTTTATTTAGCATGCAAA

>G39 Amino Acid Sequence (domain in AA coordinates: 24-90)

MPPSPKSPFISSSLKGAHEDRKFKCYRGVRKRSWGWSEIRVPKTGRRRIWLGSYDAPE
KAARAYDAALFCIRGEKGVYNFPTDKKPQLPEGSVRPLSKLDIQTATNYASSVVHVP SH
ATTLPATTPQVPSEVPASSDVSASTEITEMVDEYYLPTDATAESIFSVEDLQLDSFLMMDI
DWINNLI*

>G439 (128..967)

TATAAATCTTCGTTTCTACTTTTTTTTTCTTCCATAATATAGTCAATTCGTTTTCTTAATT
AGGGCTTCTTCTCTTTGTTTCTCCAATCTTATTAGTTTATTTATTTTGGTTATTG
TATACAAATGGCAATGGCTTTAAACATGAATGCTTACGTAGACGAGTTCATGGAAGCTCT
TGAACCATTCATGAAGGTAACCTCATCTTCTTCTACTTCGAATTCATCAAATCCAAAACC
ATTAACCTCTAATTTTATCCCTAATAATGACCAAGTCTTACCGGTATCTAACCAAAACCGG
TCCGATTGGGCTAAACCAAGCTCACTCCAACACAAATCCTCCAAATTCAGACAGAGTTACA
TCTCCGGCAAAACCAATCTCGTCGTCGCGCTGGTAGTCATCTTCTCACCCTAAACCAAC
CTCAATGAAGAAAATCGACGTAGCAACTAAACCGGTTAACTATACCGAGGCGTAAGACA
GAGGCAATGGGGTAAATGGGTAGCTGAGATTCCGGCTACCTAAAAACCGAACCCGGTTATG
GCTCGGTACGTTTCAAGACGGCTCAAGAAGCTGCATTAGCTTACGATCAAGCAGCTCATAA
GATCAGAGGAGACAACGCTCGTCTCAATTTCCAGACATTGTTCTGTCAGGACACTATAA
ACAGATATTGTCCTCCGTCTATCAACGCAAGATCGAATCCATCTGCAATAGTTCTGATCT
TCCACTGCCCTCAGATCGAGAAACAGAAACAAAACAGAGGAGGTGCTCTCTGGTTTTTCCAA
ACCGGAGAAAGAACCGGAATTTGGGGAGATATACGGATGCGGATACTCGGGCTCATCTCC
TGAGTCCGATATAACGTTGTTGGATTCTCAAGCGACTGTGTGAAAGAAGATGAGAGTTT
CTTGATGGGTTTGCACAAGTATCCTTCTTTGGAGATTGATTGGGACGCTATAGAGAACT
CTTCTGAATCCATTTTATCTTTTGGATTGATTGCTCTAAATTTGTAGAATTTTATTTTC
AGAGCTTTGTAAGGGAAGTTCTTGAATGAGAGTTGCAGAGGACTAGTGGAACCTAACTCT
GTTTTCTTTTGTAAAGTATTGTTTATAATGGGCCGTTGAATGGGCCCTTATTGATTTAAACA
GCCCAAGTTTTTAAAAAAAAAAAAAAAAAAAAA

>G439 Amino Acid Sequence (domain in AA coordinates: 110-177)

MAMALNMAYVDEFMEALEPFMKVTSSTSSNSNPKPLTPNFIPNNDQVLPVSNQTGPI
GLNQLTPTQILQIQTELHLRQNSRRRAGSHLLTAKPTSMKKIDVATKPVKLYRGVRQRQ
WGKWVAEIRLPKNRTRLWLGTFFETAQEAALAYDQAAHKIRGDNARLNFDPDIVRQGHYKQI
LSPSINAKIESICNSSDLPLPQIEKQNKTEEVLSGFSKPEKEPEFGIYCGYSGSSPES
DITLLDFSSDCVKEDSFMLGLHKYPSLEIDWDAIEKLF*

>G470 (1..2580)

ATGGCGAGTTCCGGAGTTTCAATGAAAGGTAATCGTGGAGGAGATAACTTCTCCTCCTCT
GGTTTTAGTGACCCTAAGGAGACTAGAAATGTCTCCGTGCGCCGGCGAGGGGCAAAAAGT
AATTCTACCCGATCCGCTGCGGCTGAGCGTGCTTTGGACCCTGAGGCTGCTCTTTACAGA
GAGCTATGGCAGCCTTGTGCTGGTCCGCTTGTGACGGTTCCTAGACAAGACGACCGAGTC
TTCTATTTTCTCAAGGACACATCGAGCAGGTGGAGGCTTCGACGAACCGCGGCAGAA
CAACAGATGCCCTCTATGATCTTCCGTCAAAGCTTCTCTGTCGAGTTATTAATGTAGAT
TTAAAGGCAGAGGCAGATACAGATGAAGTTTATGCGCAGATTACTCTTCTCCTGAGGCT
AATCAAGACGAGAATGCAATTGAGAAAGAAGCGCTTCTCCTCCACCTCCGAGGTTCCAG
GTGCATTGCTTCTGCAAAACCTTGACTGCATCCGACACAAGTACACATGGTGGATTTTCT
GTTCTTAGGCGACATGCGGATGAATGTCTCCACCTCTGGATATGTCTCGACAGCCTCCC
ACTCAAGAGTTAGTTGCAAAAGGATTTGCATGCAAAATGAGTGGCGATTGAGACATATATTC
CGGGGTCAACCACGAGGCAATTTGCTACAGAGTGGGTGGAGTGTGTTTGTAGCTCCAAA
AGGCTAGTTGCAGGCGATGCGTTTATATTTCTAAGGGGCGAGAATGGAGAATTAAGAGTT
GGTGTAAAGCGTGCGATGCGACAACAAGGAACGTCGCGTCTTCTGTTATATCTAGCCAT
AGCATGCATCTTGGAGTACTGGCCACCGCATGGCATGCCATTTCAACAGGGACTATGTTT
ACAGTCTACTACAAACCCAGGACGAGCCCATCTGAGTTTATTGTTCCGTTGATCAGTAT
ATGGAGTCTGTTAAGAATAACTACTCTATTGGCATGAGATTCAAAATGAGATTTGAAGGC
GAAGAGGCTCCTGAGCAGAGGTTTACTGGCACAATCGTTGGGATTGAAGAGTCTGATCCT
ACTAGGTGGCCAAAATCAAAGTGGAGATCCCTCAAGGTGAGATGGGATGAGACTTCTAGT
ATTCCTCGACCTGATAGAGTATCTCCGTGGAAGTAGAGCCAGCTCTTGCTCCTCCTGCT
TTGAGTCTCTGTTCCAATGCCTAGGCCCTAAGAGGCCAGATCAAATATAGCACCTTCATCT

CCTGACTCTTCGATGCTTACCAGAGAAGGTACAACCTAAGGCAAACATGGACCCCTTTACCA
GCAAGCGGACTTTCAAGGGTCTTGCAAGGTCAAGAATACTCGACCTTGAGGACGAAACAT
ACTGAGAGTGTAGAGTGTGATGCTCCTGAGAATTCTGTTGTCTGGCAATCTTCAGCGGAT
GATGATAAGGTTGACGTGGTTTCGGGTTCTAGAAGATATGGATCTGAGAACTGGATGTCC
TCAGCCAGGCATGAACCTACTTACACAGATTGCTCTCCGGCTTTGGGACTAACATAGAT
CCATCCCATGGTCAGCGGATACCTTTTTATGACCATTATCATCATCACCTTCTATGCCTGCA
AAGAGAATCTTGAGTGATTGAGAAGGCAAGTTCGATTATCTTGCTAACCACTGGCAGATG
ATACACTCTGTCTCTCCCTGAAGTTACATGAATCTCCTAAGGTACCTGCAGCAACTGAT
GCGTCTCTCCAAGGGCGATGCAATGTTAAATACAGCGAATATCCTGTTCTTAATGGTCTA
TCGACTGAGAATGCTGGTGGTAACTGGCCAAATACGTCCACGTGCTTTGAATTATTATGAG
GAAGTGGTCAATGCTCAAGCGCAAGCTCAGGCTAGGGAGCAAGTAACAAAACAACCCCTTC
ACGATACAAGAGGAGCAGCAAGTCAAGAGAAGGGAAGTGCAGGCTCTTTGGCATTCCT
CTGACCAACAACATGAATGGGACAGACTCAACCATGTCTCAGAGAAACAACCTTGAATGAT
GCTGCGGGGCTTACACAGATAGCATCACCAAGGTTTCAGGACCTTTCAGATCAGTCAAAA
GGGTCAAAATCAACAAACGATCATCGTGAACAGGGAAGACCATTCCAGACTAATAATCCT
CATCCGAAGGATGCTCAACGAAAACCAACTCAAGTAGGAGTTGCACAAAGGTTCAACAAG
CAGGGAATTGCACTTGGCCGTTTCAAGTGGATCTTTCAAAGTTCCAAAACATAGAGGAGTTA
GTGCTGAGCTGGACAGGCTGTTTGAAGTTCAATGGAGAGTTGATGGCTCCTAAGAAAGAT
TGGTTGATAGTTTACACAGATGAAGAGAATGATATGATGCTTGTGTTGGTGACGATCCTTGG
CAGGAGTTTGTGTTGCATGGTTTCGCAAAATCTTATATACAGAAAGAGGAAGTGAGGAAG
ATGAACCCGGGACTTTAAGCTGTAGGAGCGAGGAAGAAGCAGTTGTTGGGGAAGGATCA
GATGCAAGGACGCCAAGTCTGCATCAAACTCTTCAATGTCCAGCGCTGGGAACCTTTAA
>G470 Amino Acid Sequence (domain in AA coordinates: 61-393)
MASSEVSMKGNRGGDNFSSSGFSDPKETRNVS VAGEGQKSNSTRSAAERALDPEAALYR
ELWHACAGPLVTVPQRDDRVFYFPQGHIEQVEASTNQAAEQMPLYDLPSKLLCRVINVD
LKAEDTDEVYAQITLLPEANQDENAIEKEAPLPPPPRFQVHSFCKTLTASDSTHGGFS
VLRHHADECLPPLDMSRQPPTQELVAKDLHANEWFRHIFRQGP RRHLLQSGWSVVFSSK
RLVAGDAFIFLRGENGELRVGVRRAMRQQGNVPSVISHSMHLGVLATAWHAISTGTMF
TVYKPRTPSPSEFIVPFDQYMESVKNNYSIGMRFKMRFEGEAPEQRFTGTIVGIEESDP
TRWPKSKWRSRLKVRWDETSSIPRPDRVSPWKVEPALAPPALSPVPMRPKRPRSNIA PSS
PDSSMLTREGTTKANMDPLPASGLSRVLQGEYSTLRKHTESVECDAPENS VVWQSSAD
DDKVDVVSGSRRYGSENWMSSARHEPTYTDL LSGFGTNIDPSHGQRIPFYDHSSSPSMPA
KRILSDSEGKFDYLANQWQMIHSGLSLKLHESPKVPAATDASLQGRCNV KYSEYPVLNGL
STENAGGNWPIRPRALNYEEVVNAQAQAQAREQVTKQPFITQEETAKSREGNCR LFGIP
LTNNMNGTDSTMSQRN LNDAAGLTQIASPKVQDLSQSKGSKSTNDHREQGRPFQTNNP
HPKDAQTKTNSRSCTKVHKQGI ALGRSVDLSKFQNYEELVAELDRLFEFNGELMAPKKD
WLIVYTDEENDMMLVGDDPWQEFCCMVRKIFIYTKEEVRKMNPGT LSCRSEEEAVVGE GS
DAKDAKSASNP SLSSAGNS*

>G652 (1..606)
atgagcggaggaggagacgtgaacatgagtggtggagacagacgcaaggggaacggtgaag
tggtttgatacacagaaggggtttggtttcatcacacctagcagcgggtggtgacgatctc
ttcgttcaccagtcctccatcagatctgaaggatttcgtagcctcgcagctgaggaatct
gttgagttcgacgttgaggttgacaactccggcgtcccaaggctattgaagtgtctgga
ccgacggtgctcccgttcagggtaacagcgggtggtggtggtttcatctggtggacgcggt
ggttttgccggcgggtggtggaagaggagggggacgtggtggaggaagctacggaggaggt
tatggtggaagaggaagcgggtggcgtggaggaggtggtggtgataattcttgctttaag
tgcggtgaaccaggtcacatggcgagagaatgctctcaagggtggtggaggatacagcggga
ggcgggggtggtggaaggtacgggtctggcggcggcggaggaggaggtggtggtggctta
agctgctacagctgtggagagctctgggcactttgcaagggttgactagcgggtggtgct
cgttga

>G652 Amino Acid Sequence (domain in AA coordinates: 28-49, 137-151, 182-196)
MSGGDDVNMSGGDRRKGTWKWFD TQKGFGFITPSDGGDDL FVHQSSIRSEGF RSLAAEES
VEFDVEVDNSGRPKAIEVSGPDGAPVQGN SGGGSSSGGRGGFGGGGGGRGGGGGGGGL
YGGRGSGGRGGGGGDN SCKFCGEPGHMARECSQGGGYSGGGGGGGRYSGGGGGGGGGGL
SCYSCGESGHFARDCTSGGAR*

>G671 (61..1119)
TTCACTTGAGAACACCCCTTTGAACTCGATCAAGAAAGCTAAGTTTGAAGAATCAAGA

ATGGTGCGGACACCGTGTGCAAAGCCGAACTAGGGTTAAAGAAAGGAGCTTGGACTCCC
GAGGAAGATCAGAAGCTTCTCTTACCTTAACCGCCACGGTGAAGGTGGATGGCGAACT
CTCCCCGAAAAAGCTGGACTCAAGAGATGCGGCAAAAGCTGCAGACTGAGATGGGCCAAT
TATCTTAGACCTGACATCAAAAGAGGAGAGTTCACTGAAGACGAAGAACGTTCAATCATC
TCTCTTCACGCCCTTCACGGCAACAAATGGTCTGCTATAGCTCGTGGACTACCAGGAAGA
ACCGATAACGAGATCAAGAACTACTGGAACACTCATATCAAAAAACGTTTGATCAAGAAA
GGTATTGATCCAGTTACACACAAGGGCATAACCTCCGGTACCGACAAATCAGAAAACCTC
CCGGAGAAACAAAATGTTAATCTGACAACCTAGTGACCATGATCTTGATAATGACAAGGCG
AAGAAGAACAACAAGAATTTTGGATTATCATCGGCTAGTTTCTTGAACAAAGTAGCTAAT
AGGTTTCGGAAAGAGAATCAATCAGAGTGTTCTGTCTGAGATTATCGGAAGTGGAGGCCCA
CTTGCTTCTACTAGTCACACTACTAATACTACAACCTACAAGTGTTCCTGTTGACTCTGAA
TCAGTTAAGTCAACGAGTTCTTCCCTTCGCACCAACCTCGAATCTTCTCTGCCATGGGACC
GTTGCAACAACACCAAGTTTCATCGAACTTTGACGTTGATGGTAACGTTAATCTGACGTGT
TCTTCGTCACGTTCTCTGATTCTCCGTTAACAATCCTCTAATGTACTGCGATAATTTCT
GTTGGTAATAACAACGTTGATGATGAGGATACTATCGGGTTCTCCACATTTCTGAATGAT
GAAGATTTTATGATGTTGGAGGAGTCTTGTGTTGAAAAACACTGCGTTTATGAAAGAACTT
ACGAGGTTTCTTCACGAGGATGAAAACGACGTCGTTGATGTGACGCCGGTCTATGAACGT
CAAGACTTGTGTTGACGAAATTGATAACTATTTTGGATGAGTGAAACTCATAATCGATGAA
TCCCACGTGACCATGTCAATATGATGTCTATGGATATGTTACCTTGATGATGTTGATGGT
AATAATAATAATAATAATAGATGGTGATGATGACCATGCATGAATCATGAATGTAGTTCGTG
TTGTACATATGCTTGTGTTTTTGTGTTTTTTTTTTTTTGGTCTGAAGTGTGTTGTTTCGT
TGTAATGGATTATAAATGGTGATGTAATAATTATAATGTTAAAAA

AAAA
>G671 Amino Acid Sequence (domain in AA coordinates: 15-115)
MVRTPCCKAELGLKKGAWTPEEDQKLISYLNHRHGGWRTLPEKAGLKRCGKSCRLRWAN
YLRPDIKRGEFTEDEERSIISLHALHGNKWSAIARGLPGRDNEIKNYWNTHIKRLIKK
GIDPVTHKGITSGTDKSENLPKQNVNLTSDHDLNDKAKKNNKNFGLSSASFLNKVAN
RFGKRINQSVLSEIIGSGGPLASTSHTTNTTTTSVSVDSSESVKSTSSSFAPTSNLLCHGT
VATTPVSSNFDVGNVNLTCSSSTFSDSVNNPLMYCDNFVGNVNDDEDTIGFSTFLND
EDFMMLEESCVENTAFMKELTRFLHEDENDVVDVTPVYERQDLFDEIDNYF*

>G779 (110..712)
GACATGCATGTAAGCATTCGGTTAATTAATCGAGTCAAAGATATATATCAGTAAATACAT
ATGTGTATATTTCTGGAAAAAGAAATATATATATTGAGAAATAAGAAAAGATGAAAATGGA
AAATGGTATGTATATAAAAGAAAGGAGTGTGCGACTCTTGTGTCTCGTCCAAAAGCAGATC
CAACCAACAGCCCCAAAAGAAGCATGATGGAGCCTCAGCCTCACCATCTCCTCATGGATTG
GAACAAAGCTAATGATCTTCTCACACAAGAACACGCAGCTTTTCTCAATGATCCTCACCA
TCTCATGTTAGATCCACCTCCCGAAACCTAATTTCACTTGGACGAAGACGAAGAGTACGA
TGAAGACATGGATGCGATGAAGGAGATGCAGTACATGATCGCCGTATGCAGCCCGTAGA
CATCGACCCTGCCACGGTCCCTAAGCCGAACCGCGTAACGTAAGGATAAGCGACGATCC
TCAGACGGTGGTTGCTCGTCCGGCTCGGGAAAGGATCAGCGAGAAGATCCGAATTTCTCAA
GAGGATCGTGCCTGGTGGTGCAGATGGACACAGCTTCCATGCTCGACGAAGCCATACG
TTACACCAAGTTCTTGAAACGGCAGGTGAGGATTCTTCAGCCTCACTCTCAGATTGGAGC
TCCTATGGCTAACCCTCTTACCTTTGTTATTACCACAACTCCCAACCCTGATGAACCTAC
ACAGAAGCTCGCTAGCTAGACATTTGGTGTCATCTCTCAACCTTT

>G779 Amino Acid Sequence (domain in AA coordinates: 126-182)
MKMENGMYKKKGVCDSVSSKSRSNHSPKRSMMEPQPHLLMDWNKANDLLTQEHAAFLN
DPHLLMLDPPPETLIHLDEDEYDEMDAMKEMQYMIAVMQPVDIDPATVPKPNRRNVRI
SDDPQTIVARRRRERISEKIRILKRIVPGAKMDTASMLDEAIRYTKFLKRQVRILQPHS
QIGAPMANPSYLCYYHNSQP*

>G962 (148..1392)
CGTCTGACTCTCTACTCAACACCACTCAATTTTCTCTCTTTTTCCCTTCCATTGTTAGT
ATAAAAACCAAGCAAACCTTAATCACTTTTTCATCATCATATATCACCTTAATCCCATG
CATACACATATCTAGTCTTTTTGATATATGGCAATTGTATCCTCCACAACAAGCATCATT
CCCATGAGTAACCAAGTCAACAATAACGAAAAAGGTATAGAAGACAATGATCATAGAGGC
GGCCAAGAGAGTCATGTCCAAAATGAAGATGAAGCTGATGATCATGATCATGACATGGTC
ATGCCCGGATTTAGATTCCATCTACCGAAGAAGAACTCATAGAGTTTACCTTCGCGCA
AAAGTTGAAGGCAAACGCTTTAATGTAGAACTCATCACTTTCCTCGATCTTTATCGCTAT

GATCCTTGGGAACCTTCCTGCTATGGCGGCGATAGGAGAGAAAGAGTGGTACTTCTATGTG
CCAAGAGATCGGAAATATAGAAATGGAGATAGACCGAACCGAGTAACGACTTCAGGATAT
TGGAAAGCCACCGGAGCTGATAGGATGATCAGATCGGAGACTTCTCGGCCTATCGGATTA
AAGAAAACCCCTAGTTTTCTACTCTGGTAAAGCCCCCTAAAGGCACCTCGTACTAGTTGGATC
ATGAACGAGTATCGTCTTCCGACCATGAAACCGAGAAGTACCAAAAGGCTGAAATATCA
TTGTGCCGAGTGTAACAAAGGCCAGGAGTAGAAGATCATCCATCGGTACCACGTTCTCTC
TCCACAAGACATCATAACCATAACTCATCGACATCATCCGTTTAGCCTTAAGACAACAA
CAACACCATTTCCTCTCTAATCATTCCGACAACAACCTTAACAACAACAACAACATC
ACAATCTCGAGAAGCTCTCCACCGAATATTCCGGCGACGGCAGCACAACAACACGACC
ACAAACAGTAACCTGACGTTACCATTGCTCTAGCCAATCAAAACATATATCGTCCAATG
CCTTACGACACAAGCAACAACACATTGATAGTCTCTACGAGAAATCATCAAGACGATGAT
GAAACTGCCATTGTTGACGATCTTCAAAGACTAGTTAACTACCATAATATCAGATGGAGGT
AACATCAATCACCATAACTTTCAAATTGCTCAACAGTTTCATCATACTCAACAACAAAAT
GCTAACGCAACGCATTACAATTGGTGGCTGCGGCGACTACAGCGACAACGCTAATGCCT
CAAACTCAAGCGCGCTTAGCTATGAACATGATTCTGCGAGGAACGATTCCAAACAATGCT
TTGTGGGATATGTGGAATCCAATAGTACCAGATGGAAACAGAGATCACTATACATAATATT
CCTTTTAAGTAATTAATTAGATCATGATTATTATCCATGACAATAATTAATGCTGCTTT
CGGC

>G962 Amino Acid Sequence (domain in AA coordinates: 53-175)
MAIVSSTTSIIIPMSNQVNNNEKGIEDNDHRGGQESHVQNEDEADDHDDHDMVMPGFRFHPT
EEELIEFYLRKVEGKRFNVELITFLDLRYDPWELPAMAAIGKEWYFYVPRDRKYRNG
DRPNRVTTSGYWKATGADRMIRSETSRPIGLKTLVFPYSGKAPKGTRTSWIMNEYRLPHH
ETEKYQKAEISLCRVYKRPVEDHPSVPRSLSTRHHNHSSTSSRLALRQQQHSSSSNH
SDNNLNNNNINNNLEKLSTEYSGDGSTTTTTTNSNSDVTIALANQNIYRPMFYDTSNNTL
IVSTRNHQDDDETAIVDDLQRLVNYQISDGGNINHQYFQIAQQFHHTQQQNANANALQLV
AAATTATTLMPQTQAALAMNMI PAGTI PNNALWDMWNPIVPDGNRDHYTNIPFK*

>G977 (46..591)

CACCAAACCTCACCTGAAACCCATTTCCATTACCATTACACTAATGGCACGACCACAA
CAACGCTTTTCGAGGCGTTAGACAGAGGCATTGGGGCTCTTGGGTCTCCGAAATTCGTCAC
CCTCTCTTGAACAAAGAATCTGGCTAGGGACGTTTGAGACAGCGGAGGATGCAGCAAGG
GCCTACGACGAGGCGGCTAGGCTAATGTGTGGCCCGAGAGCTCGTACTAATTTCCCATAC
AACCTAATGCCATTCTCTCTCTTCCAAGCTTCTATCAGCAACTCTTACCGCTAAA
CTCCACAAATGCTACATGGCTTCTCTTCAAATGACCAAGCAAACGCAACACAAACGCAA
ACGCAGACCGCAAGATCACAATCCGCGACAGTGACGGTGTGACGGCTAACGAAAGTCAT
TTGAACAGAGGAGTAACGGAGACGACAGAGATCAAGTGGGAAGATGGAATGCGAATATG
CAACAGAATTTTAGGCCATTGGAGGAAGATCATATCGAGCAAATGATTGAGGAGCTGCTT
CACTACGGTTCCATTGAGCTTTGCTCTGTTTTACCAACTCAGACGCTGTGAGAAATGGCC
TTGTCTGTTTAGCGTATCTTTTCAATTTTATTTTTGTTTCCACAAAACGGCGTCGTAA
GTGATGAGAGTAGTAGTGAGAGAAGGCTAATTTCAAGACATTTTGATCTGAATTGGCCTC
TTTTGAAACACTGATTCTAGTTTCTATAAGAGCAATCGATCATATGCTATGTTATGTATA
GTATTATAAAAAAATGTTATTTTCTGATTNAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>G977 Amino Acid Sequence (domain in AA coordinates: 5-72)
MARPQQRFRGVRQRHWGSWVSEIRHPLLKTRIWLGTFFETAEDAARAYDEAARLMCGPRAR
TNFPYNPNAIPTSSSKLLSATLTAKLHKCYMASLQMTKQTQTQTQTARSQSADSDGVT
ANESHLNRGVTETTEIKWEDGNANMQNFRPLEEDHIEQMI EELLHYGSIELCSVLPTQT
L*

>G1063 (241..966)

GTAAAGAAGATGGATGGGCCACAAGTTGCTATATAAATCCTTCCACTTCTTGTGTATA
CTATTGCTTGAGTTCTGATTGGGCACAGTAGTACCATTTGCCATTCTCTCACACATACCG
TCTCTTTCTCTCATCAATCATCAATCATCCAAAAGAAAAACCCCTAAAAATTCACCTT
GTAAGCTTTTACCAGTTTCTCTCCATACCCATTTTATCAGCTTCTCCATATCTTCTCT
ATGGATTCTGACATAATGAACATGATGATGCATCAGATGGAGAAGCTTCTGAGTTTGT
AACCCTAATTCCTCTTTCTCTCTCCCGACCACAACAACACTTACCCTTTTCTCTTTAAC
TCCACTCATTACCAGTCCGATCACTCAATGACCAACGAACAGGTTTCCGCTACGGTTCC
GGTTTACTCACTAACCCTTCTTATCTCTCCCAACACAGCTTACTCTTCCGTTTTTCTT
GACAAAAGAAACAACAGTAACAACAACAATAATGGCACGACATGGCAGCTATGCGAGAG
ATGATCTTCCGTATCGCCGTGATGCAACCGATCCATATCGATCCCGAGGCGTTAAGCCA

CCGAAGAGGAGGAACGTCAGGATCTCTAAAGATCCTCAAAGCGTGGCGGCTAGGCATAGA
AGGGAGAGAATAAGCGAGAGGATTTCGGATTTTGCAACGGCTTGTTCCTGGTGGGACGAAG
ATGGATACAGCTTCGATGCTCGATGAAGCAATTCATTATGTGAAGTTTTTAAAGAAACAG
GTGCAGTCTCTGGAGGAGCAGGCGGTGGTTACTGGCGGAGGGGGAGGAGGAGGAGGAAGG
GTTTTGATCGGTGGAGTGGGAATGACGGCGGCGAGTGGTGGTGGTGGCGGCGGGGAGTG
GTTATGAAAGGGTGTGGAACAGTGGGGACTCATCAGATGGTGGGCAATGCACAGATTCTT
AGATGATGATGATTTTTAATTTTATTATTATTATTAATGTTGGAGAAAAAGAGAAAAA
TGATTCTGGAGAGGGAAGCCAAGTAATTTATGTGAGAGTCTTTAATTTAACTTTATTTTC
TTGTTTAGATAATGTGTAATGATGGTTTTTAAAGCCAAAGACTCTCCATGGTTGTTGGAG
CGAGTTTG

>G1063 Amino Acid Sequence (domain in aa coordinates: 131-182)
MDSIDIMMMMHQMEKLPEFCNPNSFFSPDHNNTPFLFNSTHYQSDHSMTNPEGFRYGS
LLTNPSSISPNNTAYSSVFLDKRNNNNNGTNMAAMREMIFRIAVMQPIHIDPEAVKP
PKRRNVRIKSDPQSVAAHRHRRERISERIRILQRLVPGGTKMDTASMLDEAIHYVKFLKKQ
VQSLLEQAVVTGGGGGGGGRVLIGGGGMTAASGGGGGGGVVMKCGTVGTHQMVGNAQIL
R*

>G1140 (67..729)

ATCCAAGATCCTCCAACCTCACAGAAAGGCAGATTCAAGAACAGTAGTGAAGGAGAGATCT
GGTAAATGGCGAGAGAGAAGATAAGGATAAAGAAGATTGATAACATAACAGCGAGACAA
GTTACTTTCTCAAAGAGAAGAAGAGGAATCTTCAAGAAAGCCGATGAACTTTCAGTTCTT
TGCGATGCTGATGTTGCTCTCATCATCTTCTCTGCCACCGGAAAGCTCTTCGAGTTCTCC
AGCTCAAGAATGAGAGACATATTGGGAAGGTATAGTCTTCATGCAAGTAACATCAACAAA
TTGATGGATCCACCTTCTACTCATCTCCGGCTTGAGAATTGTAACCTCTCCAGACTAAGT
AAGGAAGTCGAAGACAAAACCAAGCAGCTACGGAACCTGAGAGGAGAGGATCTTGATGGA
TTGAACCTAGAAGAGTTGCAGCGGCTGGAGAACTACTTGAATCCGGACTTAGCCGTGTG
TCTGAAAAGAAGGGCGAGTGTGTGATGAGCCAAATTTCTCACTTGAGAAACGGGGATCG
GAATTGGTGGATGAGAAATAAGAGACTGAGGGATAAACTAGAGACGTTGGAAGGGCAAAA
CTGACGACGCTTAAAGAGGCTTTGGAGACAGAGTCGGTGACCACAAATGTGTCAAGCTAC
GACAGTGGAACTCCCCTTGAGGATGACTCCGACACTTCCCTGAAGCTTGGGCTTCCATCT
TGGGAATGAATCTGAGAGAGAGAAAGATCCAGCAGAGTTGACTTCGATGGAAGCCACAA
ATATTAAGTCTACCTTTTCCCTTTCTTTCTTTGAATAAGTGTTGAAAAAGAATTGAGAT
GGGAAGGATGAATTCTCATTGCATTGCAGAGAAGCAAGTTTCAGATATTGTACGTGTTAT
TGGGTCTTTATAACTATTTTCTCCCCAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA

>G1140 Amino Acid Sequence (conserved domain in AA coordinates: 2-57)
MAREKIRIKKIDNITARQVTFKRRRGIFKKADELSVLCADVALIIFSATGKLFEPSSS
RMRDILGRYSLHASNINKLMDPPSTHLRLCENLSRLSKEVEDKTKQLRKLRGEDLDGLN
LEELQRLEKLLSEGLSRVSEKKGECVMSQIFLEKRGSELVDENKRLRDKLETLEKRLT
TLKEALETESVTNVSSYDSGTPLEDDSDTSLKLGLPWE*

>G1425 (43..1005)

ACTCTCTCAAACCATAAAAAATATTCTCCGATCATCATTTTAATGGAGAGTACAGATTCT
TCCGGTGGTCTCCGCCGCCGCAACCAACCTCCCTCCAGGATTCCGGTTTCATCCAACA
GACGAAGAACCTGTAATTCATTACCTCAAACGCAAAGCAGATTCTGTTCTTTTACCAGTC
GCGATCATCGCCGACGTTGATCTTTACAAATTTGATCCATGGGAACCTCCCGCGAAAGCT
TCGTTTGGAGAACAGAATGGTATTTTTTCAGTCCAAGAGATCGGAAATATCCCAACGGA
GCTAGACCTAACCAGAGCTGCGACTTCCGGTTATTGGAAAGCGACTGGTACAGATAAACCG
GTGATTTCAACCGCGCGGTGGTGGTAGTAAAAAGTGGGAGTTAAAAAGGCTCTAGTGT
TACAGTGGTAAACCACCAAAAGGAGTTAAATCAGATTGGATTATGCATGAATATCGGTTA
ACTGATAATAAACCTACTACATTTGTGACTTCCGCAACAAGAAAACTCTCTCAGGCTT
GATGATTGGGTGTTGTGTCGTATCTACAAGAAAAACAATAGTACAGCATCTAGACATCAT
CATCATCTTCATCATATTATCTAGATAATGATCATCATCGTCATGATATGATGATTGAT
GATGATCGATTCCGTCATGTTCTCTGGTCTTCACTTCCCGCGGATTTTTCTGACAAAT
AATGATCCGACGGCTATATATGATGGTGGCGGCGCGGATACGGAGGTGGAAGTTACTCG
ATGAATCATTTGTTTCGCATCTGGATCAAAGCAGGAGCAGTTGTTCCACCGGTGATGATG
ATGACTAGTCTAAATCAAGATTCGGTATTGGATCGTCGTCGTCACCTAGCAAGAGATTT
AACGGCGCGCGGCTTGGAGATTGTTGCACTTCTATGGCGGCGACGCCGTTAATGCAGAAC
CAAGGTGGGATTTACCAATTGCCTGGTTTGAATTGGTATTCTTGAAAAACAATTTACGATG
AAGAATTTTTAAATTTGTGTATATATATACGGTTTGAGTGATTAGGGGCGATTGGGGGA

TTTATTTACGGTTGATTATTATTGTAGTGTATAGAACTAAGGAGATTAAATTAAATAGA
TTGGAGGAAAAAAAAAAAAAAAAA
>G1425 Amino Acid Sequence (domain in AA coordinates: 20-173)
MESTDSSGGPPPPQPNLPPGFRFHPTDEELVIHYLKRKADSVPLPVAILIADVDLYKFDPW
ELPAKASFGEQEWYFFSPDRKYPNGARPNRAATSGYWKATGTDKPVISTGGGSSKKVGV
KKALVFYSGKPPKGVKSDWIMHEYRLTDNKPTHICDFGNKKNSLRLDDWVLCRIYKKNS
TASRHHHHLHHIHLNDHHRHDMIDDRFRHVPPGLHFPALFSDNNDPTAIYDGGGGGY
GGGYSMNHCFASGSKQEQLFPPVMMMTSLNQDSGIGSSSSPSKRFNGGGVGD CSTSMAA
TPLMQNQGGIYQLPGLNWYS*
>G1449 (105..581)
TAGACAGAGAGAAATAGAAATAGAGAGAGAGACATGAAGAGCACTCTCAATAGAGAAG
AGAAGGAAGCATGAAGCTAGCTCTGCAGCTTCAAGGTCTCATTAAAGGAGGTCTCTAACT
CTTGTTCTTCTATTTCTTCTCATCTCTGTCGACAGTACTAAACCTTCTCTTCTGAATCTT
CTGTTAATCTCTCCCTTAGTCTCTCATTTCTCTTCTACTTCTCCACAAAGAGAAGCAAGAC
AAGATTGGCCACCGATAAAGTCTAGATTAAAGAGATACACTAAAGGGTCGTCGTCTTCTTC
GTCGTGGTGATGACACTTCTCTCTTTGTTAAGGTTTATATGGAAGGTGTTCCCATTTGAA
GAAACTCGACCTTTGCGTATTCTCAGGCTACGAGAGTCTATTAGAAAATCTCTCTCACA
TGTTTCGATACTTCAATCATCTGCGGTAATCGAGATCGAAAACATCATGTTTGTGACATATG
AAGACAAGGATGGAGATTGGATGATGGTCGGAGATATTCCATGGGATATGTTTCTTGAAA
CCGTGAGAAGACTAAAGATCAGGAGACCGGAGAGGTATTAAAGCTTGGATCGGTCAAGGC
TGTGATTGCGCAGTTACGAGACGTGTAAGATTTAGGCATTGATGAAGAGACTTGAGGCGG
GACGGAGCTATTGCTGCATATTGCAACAAAGGCCTTGAAGAAGTTGGAGAATTGATTGAT
GCATATATTTATTTATATGACACCTTTGAGTGTGTTTTTCTTATAAATAAATCACAATA
TCCAAGACTTCTCTTAA
>G1449 Amino Acid Sequence (domain in AA coordinates: 48-53,74-107,122-152)
MEVSNSSCSSSSSSVDSTKPSSESSVNLSSLTFTSPQREARQDWPPIKSRLRDTLK
GRRLLRGDDTSLFVKVMEGVPIGRKLDLCVFSGYESLLENLSHMFDTSIICGNRDRKH
HVLTYEDKGDWMMVGDIPWDMFLETVRRLKITRPERY*
>G1897 (1..678)
ATGCCTTCTGAATTCAGTGAATCTCGTCGGGTTCTTAAGATTCCCCACGGCCAAGGAGGA
TCTGTTGCGATTCCGACGGATCAACAAGAGCAGCTTTCTTGCTCGCTGTGAATCAACC
AACACCAAGTTCTGTTACTACAACAATACTTCTCACAACCTCGTCATTTCTGCAAG
TCTTGTCGCGGTTACTGGACTCATGGAGGTACTCTCCGTGACATTCCCGTCGGTGGTGT
TCCCGTAAAGCTCAAACGTTCCCGGACTTATCTCTGCGCTACCACTCCGTGTGTC
GGAAGCCGGAACTTCCCTTACAAGCTACGCTGTTCTTTTCCCTCAGTCGTCTTCCAAC
GGCGGTATCACGACGGCGAAGGGAAGTGCTTCGTCGTTCTATGGCGGTTTCAGCTCTTG
ATCAACTACAACGCCCGGTGAGCAGAAATGGGCTGTTGGCGGTTTAAATGGGCCAGAT
GCTTTTGGTCTTGGGCTTGGTCACGGGTCGTTATATGAGGACGTCAGATATGGGCAAGGA
ATAACGGTCTGGCCGTTTCAAGTGCGCTACTGATGCTGCAACTACTACAAGCCACATT
GCTCAAATAACCGCCACGTGGCAGTTTGAAGGTCAAGAGAGCAAAGTCGGGTTCTGTCT
GGAGACTACGTAGCGTGA
>G1897 Amino Acid Sequence (domain in AA coordinates:34-62)
MPSEFSESRRVPKIPHGQGSVAIPTDQEQQLSCPRCESTNTKFCYNNYNFSQPRHFK
SCRRYWITHGGTLRDIPVGGVSRKSSKRSRTYSSAATTSVVGSRNFPLQATPVLPQSSSN
GGITTAAGSASSFYGGFSSLINYNAAVSRNGPGGGFNGPDAFGLGLGHGSYYEDVRYQG
ITVWPFSSGATDAATTTSHIAQIPATWQFEGQESKVG FVSGDYVA*
>G2143 (89..784)
TCTTCTTCTCTCTCCATACCTTATCTCACCAGCTTCTCCATATCTCTCAAAGAAAAACA
AACCCTATAAATCCACAAAAAGGAGGATGGATAACTCCGACATTCTAATGAACATGAT
GATGCAGCAGATGGAGAAGCTTCTGAACACTTCTCTAACTCAAACCTAACCTAATCC
CCATAACATTATGATGCTTTCTGAATCCAACACCCACCGTTCTTCTCAACCCCACTCA
TTCTCATCTCCATTGACCAAACCATGCCTCACCACCAACCCGTTTAAATTTCCGGTA
CGCCCCCTCCCGTCATCTCTCCCGGAGAAGAGAGGAGGCTGCAGCGACAACGCCAA
CATGGCGGCGATGAGAGAGATGATCTTTGAATAGCCGTGATGCAGCCTATACATATTGA
TCCGGAATCCGTAAAGCCACCAAAGAGAAAGAACGTGAGGATCTCTAAGGATCCACAGAG
CGTGGCAGCTCGGCATCGAAGGAGAGGATAAGCGAGCGGATTCGGATTCTTCAGCGGCT
TGTTCCCGGTGGGACTAAGATGGATACGGCGTCGATGCTCGATGAGGCTATCCATTACGT

TAAGTTTCTCAAGAAGCAAGTGCAGTCGCTGGAGGAACATGCGGTGGTTAACGGCGGAGG
AATGACGGCGGTGGCCGGAGGAGCACTTGCGGGTACTGTTGGTGGAGGATATGGAGGAAA
AGGGTGTGGCATTATGCGGTCTGATCATCACCAGATGCTTGGAAATGCACAGATTCTTAG
ATGATGATGATGTTGATTTTTTAAATATATATCATATGTTTATTAATATGACGGGAAAAAA
TATTATCGAGGGAGTTGAATTTAGTATCATGAACTATGAGAGCATTTTTTTTAAATGTT
TTTATCTTTCCGGGTTTCGATAATGTTTGGGATGGTTAATTAACAATTTAAAAGTCAGAC
AAGTTGGTTGTAAAGACTAAAGAATAAGCATAGTTTATCAATTTATCATTACTAAATGAA
ATAG

>G2143 Amino Acid Sequence (domain in aa coordinates: 128-179)

MDNSDILMNMMMQMEKLPFHFSNSNPNNPHNIMMLSESNTHPFFFNPTHSHLPFDQTM
PHHQPLNFRYAPSPSSSLPEKRGCCSDNANMAAMREMI FRIAVMQPIHIDPESVKPKR
KNVRISKDPQSVAARHRRERISERIRILQRLVPGGTKMDTASMLDEAIHYVKFLKKQVQS
LEEHAVVNGGGMTAVAGGALAGTVGGGYGGKGCIMRSDHHQMLGNAQILR*

>G2535 (1..1005)

ATGAACATATCAGTAAACGGACAGTCACAAGTACCTCCTGGCTTTAGGTTTACCCCAACC
GAGGAAGAGCTCTTGAAGTATTACCTCCGCAAGAAAATCTCTAACATCAAGATCGATCTC
GATGTTATTCTGACATTGATCTCAACAAGCTCGAGCCTTGGGATATTCAAGAGATGTGT
AAGATTGGAACGACGCGCGCAAAACGATTGGTACTTTTATAGCCATAAGGACAAGAAGTAT
CCCACCGGGACTAGAACCAACAGAGCCACCACGGTCGGATTTTGGAAAGCGACGGGACGT
GACAAGACCATATATACCAATGGTGATAGAATCGGGATGCGAAAGACGCTTGTCTTCTAC
AAAGGTCGAGCCCTCATGGTCAGAAATCCGATTGGATCATGCACGAATATAGACTCGAC
GAGAGTGATTAATCTCCTCGTGTGGCGATCATGACGTCAACGTAGAAACGTGTGATGTC
ATAGGAAGTGACGAAGGATGGGTGGTGTGTCGTGTTTTCAAGAAAAATAACCTTTGCAAA
AACATGATTAGTAGTACCCCGGAGTTCGGTGAAAACGCCGTCGTTCAATGAGGAGACT
ATCGAGCAACTTCTCGAAGTTATGGGGCAATCTTGTAAGGAGAGATAGTTTACAGCCCT
TTCTTAAAACTCCCTAACCTCGAATGCCATAACAACACCACCATCACGAGTTATCAGTGG
TTAATCGACGACCAAGTCAACAACCTGCCACGTGAGCAAGTTATGGATCCCAGCTTCATC
ACTAGCTGGGCGGCTTTGGATCGGCTCGTTGCCCTCACAGTTAAATGGGCCCCAAGCTCGTAT
TCAATACCAGCCGTTAATGAGACTTCACAATCACCGTATCATGGACTGAACCGGTCCGGT
TGTAATACCGGTTTAAACACCAGATTACTATATACCGGAGATTGATTTATGGAACGAGGCA
GATTTCCGCGAGAACGACATGCCACTTGTTGAACGGTAGTGGATAA

>G2535 Amino Acid Sequence (conserved domain in AA coordinates: 11-114)

MNISVNGSQVPPGFRFHPTEELLKYLRKKISNIKIDLDVIPDIDLNKLEPWDIQEMC
KIGTTPQNDWYFYSHKDKKPYPTGTRTNRATTVGFWKATGRDKTIYNGDRIGMRKTLVY
KGRAPHGQKSDWIMHEYRLDESVLISSCGDHDVNVETCDVIGSDEGWVVCVRFKNNLCK
NMISSSPASSVKTPSFNEETIEQLLEVMGQSKGEIVLDPFLKLPNLECHNNTTITSYQW
LIDDQVNNCHVSKVMDPSFITSWAALDRLVASQLNGPNSYSIPAVNETSQSPYHGLNRS
CNTGLTPDYIPEIDLWNEADFARTTCHLLNGSG*

>G2557 (94..1215)

TCGACTTCCTGTGAACATCATCTGTTTGTCTCTTCTCCGGTTTCACTTTTTCATGTCCT
GCCGTTATTACAACGAGGATTGTGTTTGATCCGATGGAAGGATTGGAATCTGTGTACGCT
CAAGCTATGTATGGAATGACACGAGAGAGCAAAATCATGGAGCATCAAGGATCAGATTG
ATTTGGGGAGGAAATGAGCTAATGGCTCGAGAACTCTGTCTTCTTCTTCTTATCACCAC
CAACTCATTAAATCCGAATCTTAGCAGCTGTTTCATGTCTGATCTTGGAGTCTTAGGTGAG
ATTCAACAGCAGCAACATGTTGGCAACAGAGCTAGCTCGATAGATCCATCATCACTCGAT
TGTTTGTATTCTGCGACGTCCAATAGCAACAACACCTCGACGGAGGACGATGAAGGAATA
TCTGTGCTTTTCTCAGATTGTCAGACTCTTTGGAGCTTGGTGGAGTCTCATCTGCAGAG
TCTGAGAACAGAGAGATCACTACTGAGACGACAACAACGATAAAGCCTAAGCCTTTGAAG
AGAAACAGAGGAGGAGATGGAGGAAGTACTGAGACTACAACAACAACAACAAACCTAAG
TCTTTGAAGAGAAACAGAGGAGACGAGACAGGAAGTCACTTTAGTCTTGTTCATCCTCAA
GATGATTCGGAGAAAGGAGGTTTCAAGCTTATATACGATGAGAATCAATCGAAATCAAAG
AAACCAAGAACAGAGAAAGAACGAGGCGGTTCTTGAACATTAGTTTCCAACATTCAACT
TGTTTGTCTGCAATGTGCGAGCCCGATGCTGAGGCGATTGCACAAATGAAGGAGATGATA
TACAGAGCGGCTGCATTTAGACCGGTGAATTTCCGGTTAGAGATTGTGGAGAAGCCTAAG
AGGAAGAACGTCAAGATATCGACGGATCCTCAAACGGTTGCAGCGAGACAGAGAAGGGAG
AGGATAAGTGAGAAGATTAGGGTTTACAAAATTGGTTCCAGGTGGGACGAAGATGGAT
ACTGCATCAATGCTTGATGAAGCTGCTAATTATCTCAAGTTCTTAGAGCACAAGTAAAA

GCTTTAGAAAACCTTGAGACCCAAGCTTGACCAAACCAATCTCTCTTTCTCTCTGCTCCT
 ACATCGTTTCCATTATTCACCCCATCTTTCTTCCATTGCAAAATCCTAATCAAATCCAT
 CATCCAGAGTGTGACAGATTATAAACTTTTGAGTTTCATCATCATCAACAGAATCATGG
 CGTCTTGATTGTTTTAGCAGTTCTCAAGAAAGGCAACTTCTGTGACAAGGGTGGTGTCTGG
 GCAGTGTGTTTACACTTTCAGTCTTTGTTTTGCATTTCTTTTATATAAAGTTTGTAT
 TTTATATAGAATCTGTGGAATTCGAGGGTTGAAATATTGTGAAAAACAGAGCCGCAAGAG
 GTTAATTACAGTCTCTGCAATATTTTCAACCTTTTATTACTTTATTAGAGTAAAGATAGC
 GT

>G2557 Amino Acid Sequence (domain in aa coordinates: 278-328)

MEGLESVYAQAMYGMTRESKIMEHQSDLIWGGNELMARELCSSSSYHHQLINPNLSSCF
 MSDLGVLGEIQQQHVGNRASSIDPSSLDCLLSATSNNSNTSTEDDEGISVLPSCQTLW
 SFGGVSSAESENREITTTETTTTIKPKPLKRNRRGGDGGTTETTTTTPKPSLKRNRRDETG
 SHFSLVHPQDDSEKGGFKLIYDENQSKSKPRTEKERGGSSNISFQHSTCLSDNVEPDAE
 AIAQMKEMIYRAAARFVNFGLIVEKPKRKNVKISTDPQTVAARQRRERISEKIRVLQT
 LVPGGTKMDTASMLDEAANYLKFRLAQVKALENLRPKLDQTNLSFSSAPTSFPLFHPSPFL
 PLQNPNIHHEC*

>G259 (52..786)

GAGATCTTCTACTACTGTGTTTTCTTCAAGAATAATAATTTTCGTTTATATATGGAAGAT
 GCTGGTGAACATTTACGGTGTAAACGATAACGTTAACGACGAGGAGCGTTTGCCATTGGAG
 TTTATGATCGGAACTCAACATCCACGGCGGAGCTACAGCCGCCTCCACCGTTCTTGGTA
 AAGACATACAAAGTGGTGGAGGATCCGACGACGGACGGGGTTATATCTTGGAAACGAATAC
 GGAACCTGGTTTCGTCTGTGGCAGCCGCGAGAATTCGCTAGAGATCTGTTACCAACACTT
 TTCAAGCATTGCAACTTCTCTAGCTTCGTTCCGACGCTCAATACTTACGGTTTTTCGAAAA
 GTAACGACGATAAGATGGGAATTTAGTAATGAGATGTTTCGAAAGGGGCAAAGAGAGCTT
 ATGAGCAATATCCGAAGAAGGAAGAGCCAAACATTGGTCACACAACAAGTCTAATCACCAG
 GTTGTACCAACAACAACGATGGTGAATCAAGAAGGTCATCAACGGATTGGGATTGATCAT
 CACCATGAGGATCAACAGTCTTCCGCCACTTCATCCTCTTTCGTATACACTGCATTACTC
 GACGAAAAACAATGCTTGAAGAATGAAAACGAGTTATTAAGCTGCGAACTTGGGAAAAACC
 AAGAAGAAATGCAAGCAGCTTATGGAGTTGGTGGAGAGATACAGAGGAGAAGACGAAGAT
 GCAACTGATGAAAGTGAATGAAGAAGATGAAGGGCTTAAGTTGTTCCGAGTAAACTT
 GAATGAACTAGATTGCTAGATTGATATTCGTAATATACCAGTTTCTTCATATTCTTAGA
 AGTTTTGCATAACTATATATAGTACTCTTTTAAGACATGCAAGATCAGAACATATG

>G259 Amino Acid Sequence (domain in AA coordinates: 27-131)

MEDAGEHLRCNDNVNDEERLPLEFMIGNSTSTAELOPPPPFLVKTYKVVEDPTTDGVISW
 NEYGTGFVWQPAEFARDLLPTLFKHCNFSFVRQLNTYGRKVTITIRWEFSNEMFRKQ
 RELMSNIRRRKSQHWSHNSNHQVVPVPTTMTVNQEGHQIRIGIDHHHEDQQSSATSSSFVYT
 ALLDENKCLKNENELLSCELGKTKKKCKQLMELVERYRGEDEDATDESDDDEDEGLKLF
 VKLE*

>G353 (82..570)

ACCAAACCTCAAAAAACACAAACCACAAGAGGATCATTTTCATTTTTTATTGTTTCGTTTTA
 ATCATCATCATCAGAAGAAAAATGGTTGCGATATCGGAGATCAAGTCGACGGTGGATGTC
 ACGGCGGCGAATTTGTTTGATGCTTTTATCTAGAGTTGGACAAGAAAACGTTGACGGTGGC
 GATCAAAAACGCGTTTTTACATGTAAAACGTGTTTGAAGCAGTTTCATTCGTTCCAAGCC
 TTAGGAGGTCACCGTGCGAGTCACAAGAAGCCTAACACGACGCTTTGTCGTCGATTG
 ATGAAGAAGGTGAAAACGTCGTCGATCCTTGTCCCATATGTGGAGTGGAGTTTCCGATG
 GGACAAGCTTTGGGAGGACACATGAGGAGACACAGGAACGAGAGTGGGGCTGCTGGTGGC
 GCGTTGGTTACACGCGCTTTGTTGCCGAGCCACGGTGACTACGTTGAAGAAATCTAGC
 AGTGGGAAGAGAGTGGCTTGTGATCTGAGTCTAGGGATGGTGGACAATTGAATCTC
 AAGTTGGAGCTTGAAGAACAGTTTATTGATTTTATTTATTTCTCTTAAATTTTCTGAAT
 ATATTGTTTTCTCTCATTCTTTGAATTTTCTTAATATTCTAGATTATACATACATCCGC
 AGATTAGGAAACTTTCATAGAGTGAATCTTTCTTTCTGTAAAAATATATTTTACTTG
 TAGCAAA

>G353 Amino Acid Sequence (domain in aa coordinates: 41-61, 84-104)

MVAISEIKSTVDVTAANCLMLLSRVGQENVDDGQKRVFTCKTCLKQFHSFQALGGHRAS
 HKKPNNDALSSGLMKVKVTSHPICGVEFPMGQALGGHMRRHRNESGAAGGALVTRAL
 LPEPTVITLKKSSSGKRVACLDSLGMVDNLNLKLELGRVY*

>G354 (27..533)

CCTAGAAGTCACTAAGTCGATTCAAAATGGTTGCGAGAAGTGAGGAAATTGTGATAGTGG
AAGAAGATACGACTGCGAAATGTTTGATGTTTATCAAGAGTCGGAGAATGCGGCGGCG
GCTCGGGGGGAGATGAACGTGTTTTCCGATGCAAGACTTGTCTTAAAGAGTTCTCATCGT
TTCAAGCTTTGGGAGGTTCATCGTCAAGCCACAAGAACTTATCAACAGTGACAATCCAT
CACTTCTTGGATCCTTGTCCAACAAGAACTAAAACGTCTCATCCTTGTCCGATATGTG
GAGTGAAGTTTCCGATGGGACAGCTCTTGGTGGTCACATGAGGAGACATAGGAACGAGA
AAGTCTCAGGCTCGTTGGTTACACGTTCTTTTCTACCGGAGACGACGACGGTGACGGCTT
TGAAGAAATTTAGTAGTGGGAAGAGAGTGGCTTGTGGATTGGAGTTAGATTTCGATGG
AGAGTTTGGTCAATTGGAAGTTGGAGTTGGGAAGAAGGATTCTTGGAGTTAAGTTTTTG
GGTTGTATACAGTTTACATGATTTTGTATCTTTGTTGATCCAATTATCGTACCGATCG
ATGTGAATATTATTTTGATACAATAAAA

>G354 Amino Acid Sequence (domain in aa coordinates: 42-62, 88-109)

MVARSEEIVVEEDTTAKCLMLLSRVGECGGGCGGDERVFRCKTCLKEFSSFQALGGHRA
SHKKLINSNDNPSLLGSLSNKKTKTSHPCPICGVKPFPMQALGGHMRRHRNEKVSGLVTR
SFLPETTTVTALKKFSSGKRVACLDLDLSMESLVNWKLELGRITISWS*

>G638 (86..1861)

GAATTAAGAGTTTAACTTTTACCTTTTTTCCCTTCACTATCGATAATTGATCTTCTCT
TTCGGCTGAATATAAATCTGAAAAATGGATCAAGATCAGCATCCTCAGTACGGTATACC
GGAGCTCCGGCAGCTCATGAAAGGCGGAGGAAGGACGACTACTACAACACCGTCTACTTC
TTCTCATTTTCCCTCTGATTTCTTTCGGTTTAACTTGTCTCCGGTGCAGCCACCGCCACA
CCGTCTTCATCAGTTCCTACTGATCAAGATATGGGTTTCTTGCCACGTGGCATAACATGG
ATTGGGTGGAGGTTCTTCAACGGCTGGAAATAACAGTAACTTAAACGCGAGTACTAGTGG
TGGAGGAGTTGGGTTTGTGGGTTTCTTGACGGTGGTGGTTTCCGGCAGCGGAGTAGGAGG
AGACGGTGGAGGAAGTGAAGGTGGCCGAGACAAGAAACCTAACTCTGTTGGAAATTAG
ATCTCGTCTTGATCATAAATTCAAAGAAGCTAATCATAAAGGACCTCTTGGGATGAAGT
TTCTAGGATTATGTCCGAGGAACATGGATACCAAAGGAGTGGGAAGAAATGCAGAGAGAA
GTTTGAGAATCTGTACAAATACTATAGTAAGACTAAAGAAGGCGAAGCCGGAAGACAAGA
CGGAAACATCACAGATTTTTCCGCCAGCTCCAAGCGCTATACGGGGATTCTAATAACTT
GGTTTCTTGTCCTCAATCATAACACGCGAGTTCATGAGCAGTGCTCTTCATGGTTTCCATAC
TCAAAACCTATGAACTGTGCTACAAACAGTCCAACATCCATAACGTTGATAGTGTTC
TGTTTTCATCAAAGCCTTAGTCTTTCTAACAACTACAACCTCCTCCGAGCTTGAGCTGAT
GACTTCTCTTTCGGAAGGGAATGATTCTAGTAGTAGAAGGAAAAAGAGGAGTTGGAAGC
GAAGATAAAGGAGTTTATTGATACGAACATGAAAAGGTTGATAGAGAGGCAAGATGTTTG
GCTTGAGAAGTTGACAAAGGTTATTGAAGACAAAGAGGAACAACGGATGATGAAAGAAGA
GGAATGGAGGAAGATTGAAGCTGCAAGGATTGATAAAGAGCATTGTTTTGGGCTAAAGA
GAGGCGAGGATGGAAGCTAGGGATGTTGCGGTGATTGAGGCAATTGCAATATTGACAGG
AAAGCCATTGATAAAGCCGCTGTGTTTCATCCCGGAAGAGAGGACAAATGGTAATAATGA
GATCCGAAACAATAGTGAGACACAGAATGAGAATGGAAGCGATCAAACGATGACTAACAA
TGTTTGTGTTAAAGGAAGTAGTAGTGTGCTGGGCTGAGCAAGAGATTTTAAAGCTTATGGA
GATAAGAACGAGCATGGACTCGACCTTTCAAGAGATATTAGGAGGGTGTCTCGGATGAGTT
TCTATGGGAGGAAATCGCAGCGAAGTTGATTAGTTAGGGTTTGTATCAGAGAAGTGCCTT
ATTATGCAAGGAAAAGTGGGAATGGATAAGCAATGGAATGAGGAAAGAAAAGAAGCAAAT
CAACAAGAAAAGAAAGGATAATTCGTCCAGCTGCGGCGGTACTACCCGAGAAACGAAGA
AAATCCAATCTACAATAATCGAGAAAGTGGATATAATGATAATGATCCGCATCAAATCAA
CGAACAAGGCAATGTAGGTTCTTCAACATCAAACGCAAACGCAAACGCAAACGTAACCAC
TGGAAATCCGAGCGGTGCAATGGCTGCTAGTACAAACTGCTTCCCGTTCTTCATGGGAGA
TGGAGATCAGAATTTGTGGGAGAGTTATGCTTTGAGGCTCAGTAAAGAAGAGAATCAGTA
AGTAATTTCTCTTAATGAAGAAGAAGAAGTAATCATGTGGTTAACTAATTCTTTTGTAGT
TAGCTATATATGAGATAAACCTTGACTTAGCTATTATATGTACATGTGCTTAGAATTA
AGAAATATTGTTGGGGCTTAACGAATTATATATCAGCATATATAAGATGAGAGTCTAAG
AATTATATCAAATTAGGCTTTAACCACGTACGATTATATATTATGTTTTTCATGTATTTA
TTCTGTAAAGACTTTTTAATATCAATCTTCTCTAAA

>G638 Amino Acid Sequence (domain in AA coordinates: 119-206)

MDQDQHPQYGIPELRQLMKGGGRTTTTTPTSSHFPSDFFGFNLPVQPPPHRLHQFTTD
QDMGFLPRGIHGLGGSSSTAGNNSNLNASTSGGGVGFSGFLDGGGFGSGVGGDGGGTGRW
PRQETLTLLLEIRSRDLHKFKKANHKGPLWDEVSRIMSEEHGYQSRGKKCREKFENLYKYY
SKTKEGBAGRQDQKHHRFRQLQALYGDSNNLVSCPNHNTQFMSSALHGFHTQNPMMNVAT

TTSNIHNVDSVHGFHQSLSLSNYNSSSELEMTSSSEGNDSRRKKRSWKAKIKEFIDT
 NMKRLIERQDVWLEKLTKEVIEDKEEQRMKEEWRKIEAARIDKEHLFWAKERARMEARD
 VAVIEALQYLTKGKPLIKPLCSSPEERTNGNNEIRNNSETQNEGSDQTMNNVCVKGSSS
 CWGEQEILKLMEIRTSMDSTFQEILGGCSDEFLWEEIAAKLIQLGFDQRSALLCKEKWEW
 ISNGMRKEKKQINKRKDNSSSCGVYPRNEENPIYNNRESGYNDNDPHQINEQGNVGSS
 TSNANANANVTGPNPSGAMAASNCFFPFMGDGDQNLWESYGLRLSKEENQ*

>G869 (428..1402)

AGGAACAGTGAAAGGTTCCGGTTTTTTGGGTTTCGATCTGATAATCAACAAGAAAAAGGG
 TTTGATTTATGTCGGCTGGGTTTGAATCGACTGTGATTTTGTCTTTGATTTCATATCTCTT
 CTCCGATTTTCATCATCATCTTCCCCATCATCGTCGTTTGAATCTTGTCTTCTCAACG
 CTCTTCACTTCTGTCTGTATAAGCAGAGGCTTGTCTTGGAGACTCCTTCTCTTCCATGC
 GCTTAAGACCCAAAAGGACTTGTCTAGTGTGAAGTCTTTGGGGTTTTTACATAAAGC
 AGCAAAAGTTTTCTTTTTTCATAGTTCGCTGAGAGTTTGAAGTTTGATACCAAAAAAGT
 TTTGACCTTTTAGAGTGATTTTTTGTCTTCTGTCTTCTGGGTATTTTTGAGGAGTGGG
 TTTAACAATGGTTGCGATTAGAAAGGAACAGTCTTTGAGTGGTGTAGTAGCGAGATTAA
 GAAGAGAGCTAAGAGAAACACTCTATCGTCCCTTCTCAAGAAACCCAACTTTGAGGAA
 AGTCCGTATTATTGTGAATGATCCTTATGCTACTGATGATTCCTCTAGTGATGAGGAAGA
 GCTTAAGGTTCTTAAAGCAAGGAAATGAACAGTATCGTTCTGAGATTAACTTTCTCTTC
 TATGGAAGTTTTCTGAACAGCCTTCTGAGAGTCTTCTCAGGACAGTACTAAAACGTATGG
 CAAGATAGCTGTGTGCTGCTTCTCCTGCTGTTCTTAGGAAGAAGCCTGTTGGTGTTAGGCA
 AAGGAAATGGGGGAAATGGGCTGTGAGATTAGAGATCCTATTAAAGAAACTAGGACTTG
 GTTGGGTACTTTTTGATACTCTTGAAGAAGCTGCTAAAGCTTATGATGCTAAGAAGCTTGA
 GTTTGATGCTATTGTTGCTGGAATGTGTCCACTACTAAACGTGATGTTTCTTCATCTGA
 GACTAGCCAATGCTCTCGTTCTTACCTGTTGTTCTGTTGAGCAAGATGACACTTCTGC
 ATCAGTCTCTCATTGTGTCAACAACCTGATGACGTCFCGACCGTTGCTCCAACGTCTCC
 AACTCCAAATGTTCTGCTGGTGGAAACAAGGAAACGTTGTTGATTTCTGACTTTACTAA
 TCTACAGATCCCTGATTTTGGTTTTCTTGGCAGAGGAGCAACAAGACCTAGACTTCGATTG
 TTTCTCGCGGATGATCAGTTTGTGATTTTGGCTTGCTTGATGACATTCAAGGATTCTGA
 AGATAACGGTCCAAGTGCCTTACCAGATTTGACTTTGCGGATGTTGAAGATCTTCAGCT
 AGCTGACTCTAGTTTCCGTTTCTTGTATCAACTTGCTCCTATCAACATCTCTTGCCCAT
 AAAAGTTTTGCAGCTTCATAGGATCTTGCTTAGTAATGTTAAGTGAGAAGAGTGTTTG
 TTTTTTCGTTTATGCTTTAGTAATTTAAGACATACAAAGTGTGTGTTCCGGATTGTAGT
 AAGATCTTAAGACATAAAGCCGGTTTTTGCAATTAGGAATCGAGTTTAATGAAGTTTA
 GTTTATGTTTG

>G869 Amino Acid Sequence (domain in AA coordinates: 109-177)

MVAIRKEQSLSGVSSEIKKRAKRNLTSSLPQETQPLRKVRRIIVNDPYATDDSSSDEEELK
 VPKPRKMKRIVREINFPSMEVSEQPSSESSQDSTKTDGKIAVSASPAVPRKKPVGVRQRK
 WGWAAEIRDPIKTRTWLGTFTDLEAAKAYDAKKLEFDAIVAGNVSTTKRDVSSSETS
 QCSRSSPVVPVEQDDTSASALTCVNNPDDVSTVAPTPTPNVPAGNKETLFDFTNLQ
 IPDFGFLAEEQQDLDFDCFLADDQFDDFGLLDDIQGFEDNGPSALPDFDFADVEDLQLAD
 SSFGFLDQLAPINISCLPKSFAAS*

>G1645 (25..1104)

CGTCGACCTCCCAACACTAACTCCATGTTTATAACGGAAAAACAAGTGTGGATGGATGAG
 ATCGTCGCAAGAAGAGCTTCTTCTTCTTGGGACTTCCCTTTCAACGACATTAATATTCAT
 CAGCATCATCATCGTCACTGCAACACAAGTCATGAGTTTGAATCTTGAAGAGTCTCTT
 GGAGATGTAGCGGTTACGAAGAAGAGAGTAATAATAATAACCTAATTTTCAGTAACAGC
 GAGAGTGGTAAGAAGGAGACAACAGATAGTGGTCACTTGGTCTCTCGTCTCTTCAAAA
 CCATCGGTCTTGGGGAGAGGACATTGGAGACCAGCTGAAGATGTTAAACTCAAAGAGCTT
 GTCTCCATTTACGGCCCAAAAACCTGGAACCTCATAGCTGAAAAGCTTCAAGGAAGATCT
 GGGAGAGCTGTAGACTACGATGGTTTAACCAATTGGACCCGAGGATAAACCGAAGAGCT
 TTCACAGAAGAAGAAGAGGAGAGGCTGATGCAAGCACATAGGCTTTATGGTAACAAATGG
 GCAATGATTGCGAGGCTTTTCCCTGGTAGAACTGATAATTCACTGAAGAACCATTGGCAT
 GTTGTCTAGGCTCGTAAGTATAGAGAACACTCTTCTGCTTACCGTAGGAGAAAGCTTATG
 AGTAATAATCCACTTAAACCTCACCTCACCAATAATCATCATCCTAACCTAACCTAAT
 TACCACTCTTTTATCTCCACTAATCATTACTTCGCTCAGCCTTTCCCCGAGTTTAATTTG
 ACTCATCACTGGTTAATAATGCCCTATCACGAGTGACCATAACCAGCTTGTGTTGCCT
 TTCCATTGCTTTCAAGTTATGAGAACAAATGAACCTCCGATGGTTGTGAGTATGTTGGC

AACCAAATGATGGTCGGCGATAACGTTGGTGCCACGTCAGACGCGTTATGCAATATTCCG
CACATTGACCCTAGTAACCAAGAGAAACCGGAGCCAAATGATGCAATGCATTGGATCGGA
ATGGACGCGGTAGATGAGGAGGTGTTGAAAAGGCTAAGCAGCAACCACATTTTTTCGAT
TTTCTTGGCTTGGGGACGGCGTGAATGTTGAACAAATTGGTGTTAATCAGATAACGACAG
TGGC

>G1645 Amino Acid Sequence (domain in AA coordinates: 90-210)

MFITEKQVWMDEIVARRASSSWDFPNDINIHQHHHRHCNTSHEFEILKSPLGDVAVHEE
ESNNNNPNFNSNESGKKETDSDGQSWSSSSSKPSVLGRGHRPAEDVKLKLVSIIYGPQN
WNLIAEKLQGRSGKSCRLRWFNQLDPRINRRAFTEEEERLMQAHRLYGNKWAMIARLFP
GRTDNSVKNHWHVVMARKYREHSSAYRRRKLMSNNPLKPHLTNNHHPNPNPNYHSFISTN
HYFAQPFPEFNLTHLVLNAPITSDHNQLVLPFHCFQGYENNEPPMVVSMFGNQMMVGDN
VGATSDALCNIPHIDPSNQEKPEPNDAMHWIGMDAVDEEVFEKAKQQPHFFDFLGLGTA*

>G1038 (240..1574)

GCTCGTTTTCAAATTAAAAACAGGGAGAAATTTGGAAATTCAGTACGACGGGAGATAAA
ACCTAACATACGCCATGGTGACCGTTATCTAAACTACGCCAAAATATTTGAAGTGTCGTC
GTTTCATAATAAAACGCAACAAAAACCCACTCCACTTTCTCCTTTCCAAAAAAGAAC
TCTCGCCACTTTCTCTGCTCTTTCTTTCTCTCTCTCTTTCTTGTTTTCGCCGGCGATCA
TGGAGAAAAGCGGCTTCTCTCCGTCGGTCTAAGGGTTCTTGTCGTAGACGATGATCCAA
CTTGGCTCAAGATTCTCGAGAAAATGCTCAAGAAGTGTCTTACGAAGTAACGACCTGTG
GATTAGCTAGAGAGGCTTTGAGGTTGCTGAGGGAGCGTAAAGATGGATATGATATCGTGA
TCAGCGATGTGAACATGCCTGACATGGATGGTTTCAAGCTTCTTGAGCATGTTGGTCTTG
AATTAGACCTCCCTGTAATAATGATGTCGGTGGACGGCGAAACAAGCCGAGTGATGAAGG
GACTGCACACGGGAGCTTGTGATTACCTCTTGAAGCCGATAAGAATGAAGGAGTTAAAGA
TTATATGGCAACATGTTCTGAGAAAAGAGCTTCAAGAAGTGAGAGATATCGAAGGCTGTG
GATACGAAGGAGGAGCGGATTGGATCACTCGATACGATGAAGCACATTTTCTTGGAGGTG
GTGAAGATGTTTTCTTTGGGAAAAGAGAAAAGACTTTGACTTTGAGAAGAAGCTTCTTC
AAGATGAGAGTGATGTCATCATCTTCTTCTTCAAGAAAGCTAGAGTTGTTTGGTCTTTTG
AGCTTCATCATAAGTTTGTCAACGCCGTTAACCAAATCGGATGCGATCACAAAGCTGGTC
CCAAGAAGATATTGGATCTCATGAATGTTCCATGGCTCACTAGAGAAAATGTTGCAAGCC
ACCTTCAGAAATATAGACTTTACCTGAGCAGATTAGAGAAAGGAAAGGAGCTCAAGTGTT
ATTCAGGTGGCGTGAAGAATGCGGATTCTCTCCAAAGATGTCGAAGTGAATTCAGGCT
ACCAAAGCCCTGGGAGGAGCAGCTATGATTCTCTGGAGGAAATCTCTGATCCAAAAAG
CAACAGAGATTGATCCAAAGCCACTTGCTTCAGCTTCTTTGTCTGACCCCAACACCGATG
TGATCATGCCTCCGAAAAACAAAAAGACGCGTATAGGATTGATCCTCCATTTCTCCT
CTGCGTTTGACTCTCTGCTTCTTGAATGATGTTCCAGAGGTCCTTGAATCGAAGCCGG
TTCTGTATGAGAATAGCTTTCTCCAGCAACAACCATTGCCAAGTCAAAGTTCTTATGTTG
CAATTTCTGCACCATCTCTCATGGAGGAGGAAATGAAGCCTCCTTATGAGACACCAGCAG
GAGGCAGTAGTGTGAATGCAGATGAGTTTCTCATGCCACAAGACAAGATCCCTACTGTAA
CCCTTCAAGATTGGATCCCTCTGCCATGAAGCTGCAGGAGTTCAACACAGAAGGCGATT
CTGAAGAAGCTTGAACCTGGGAACTTCCAGAATCACATCATTCTGTTTCTTTAGACACTG
ACTTAGACTTGACTTGGCTTCAAGGCGAGCGTTTCTTGCAAACACCGACTCCAGTTTCAA
GATACAGTAGTAGCCCATCACTCCTATCTGAGCTCCCAGCCACCTTAATTGGTATGGAA
ATGAGCGGCTGCCTGACCCTGACGAGTATTCCTTCATGGTAGACCAAGGTTTATTCTAT
CTTAACCTTGTTCCAATAACTTCTTTTCGTATATTGGTTGGTGTAAATGCAGAAAGATTTT
GTGGGTATACCTGAAAATAATCTTGCTTTCCCAAGAACCTTCCATGATCGGATGCATTGT
ACAATAATCCAGAGTGTCTGATAGGCTAATTACACCAAACAGGTTGATGACAGTGATAAGG
CCACATGTTTACACACCGTCGCTTAAGATCTTTACTGTCACTTGAAGGAAA

>G1038 Amino Acid Sequence (domain in AA coordinates: 198-247)

MEKSGFSPVGLRVLVDDPTWLKILEKMLKKCSYEVTTCLAREALRLLRERKDGDIV
ISDVNMPDMDGFKLLEHVGLLELDLPVIMMSVDGETSRVMKGVHTGACDYLLKPIRMKELK
IIWQHVLRLKKLQEVVDIEGCGYEGGADWITRYDEAHFLGGGEDVSFGKKRKDFDFEKKLL
QDESPPSSSSSKARVVWSFELHHKFVNNAVNIQCDHKAGPKKILDLNVPWLRENVAS
HLQKYRLYLSRLEKKGELKCYSGGVKNADSSPKDVEVNSGYQSPGRSSYVFGGNSLIQK
ATEIDPKPLASLSDPNTDVIIMPPKTKKTRIGFDPPISSAFDSLPLPNDVPEVLESKP
VLYENSFLQQQLPLPSQSSYVAISAPSLMEEEMKPPYETPAGGSSVNADEFMLPQDKIPTV
TLQDLDPKAMKLQEFNTEGDSEEA*

>G1073 (62..874)

ctctctctagcgctcactctctctctctcattggtcggtagaataaaggccaaggaagggatca
gttttaagttttgtttcattctttttgtagctggagaaaagagtttttgaaaaataaaaac
aꝥaaaaaatgccgattaggcaaatgaaagatagctctgagactcacttagttatcaaaa
cccaacctttaagcaccacaatccaaaaaccggttcaaaaccggtaaaaatcctcctcctt
ctccttctccggtgacggtgactactccggcgacggttactcagagtcaagcttcttcac
cttcaccaccgctcaaagaatcgtagccggaggagaaaccgctggtggaagaaaatctgatc
aaggagatgtttgtatgagacctagctctcgtcctcgtaaacccgccaccgccaagtcaaa
ccactctctccgcgctctccgtcgccaccgcgggtgagattgtcgtctgtaaatcatcaga
tgcagatgggtgttcgtaaaaaactcaaaacttgcctcaagacctggatttggaacacttg
gaactaaatgcattgtttaagctaaccactttctcgtctgatttgcctaccaggatttga
atcagtatgatgttacaattactcctgaagtgatcaaaagagtgttaacagagctataa
ttgctgagttagttagactttacaaagagtctgatctcgggaggagacttccggcttacg
atggccggaagagtctttacactgctggagaactctctttacttggaaggagttcagtg
ttaagattgttgatgaagatgacggtatcatcaatggccctaaaagggaagagatcatata
agggtggcaatcaagtttgttgccagggcaaatatgcatcacttaggcgagtttctagctg
gtaaacgggcagattgtccgcaagagggtgcagattcttgatattgtactcaggaggt
tgtcgggttaagagggttttgtccggttggaaagattctcttctcgcctgatattaaaacac
cgcagcgactcgggtgaagggttagagtcaggtgtgggttttaccagagatttagaccaa
ctcaaatgggtttatcactaaatatcgatatggcttcagctgcattcatcgagcctcttc
cagtgatagagttttagcacagcttcttggaaaggatgtcttgtcgaagccatttgtcgg
attctgactcgcgtcaagattagaagggtcttagaggagtgaagtagagggttactcaca
gacgcaatgttaagaaggaaaatccgtgttgcggttttaacaactcaaccaacaagagagc
taatgtttccagtagatgagaactgtactgaagtcagttatttgatatttccaagaga
tgtatggattcacgatccagcacacgcatttgccatgtctcgaagtggaaacccaaaaa
aggcaagctatttgcgatggaggcatgcaaaattgtcgagggaacaacggttacagaaaa
ggttgaatgagaagcagattactgtctctctgaaagttacatgccaaaggggccgagggac
agagaaacgatattttgcggactgtccaacacacgcataatgatcaagatccatatgcaa
aggagtttggcattgaacataagcgaaggttagcttctgttgaaagctcgtattcttccag
ctccatggcttaagtatcacgagaacgggaaagaaaaagattgtctcccgcaagttgggtc
agtggaatatgatgaacaagaaaatgataacgggagtgactgtgagcagatgggcctgtg
ttaacttctcacgcagcggttcaagaaaaacgtctcgtggatttgtaatgaactatggtc
agatgtgtgaagcttcaggcatggagtttaatccagaaccggtgataccaatatatagtg
cgaggcccgatcaagtcgagaaagctctaaagcatgtttatcacacttcaatgaacaaaa

cctaaaggcaagaggttagagcttctgctggcaatattacctgataacaacgggttcacttt
 atggtgatcttaagagaatctgtgaaaccgagcttggtttgatatctcaatggtgtctca
 caaaacatgtgttcaagatttagcaaacagtatctggcagatgtatcccttaaaatcaacg
 taaagatgggaggaaggaacacagttctagtagacgccataagctgtagaattccactgg
 ttagcgatataccgacaatcatttttggcgcagacgtgactcaccagagaacggggaag
 agtcaagcccttcaatcgctgctgtgtgtgtcttcaagactggcctgaagtgacaaaat
 atgcggttttagtttggctcaagctcacaggcaagaacttatacaagatttgtataaaa
 catggcaagatcctgttcgcggtactgttagtgcggtatgatcagggaccttcttatct
 catttagaaaagcaacagggcaaaaaccgcttcgaattatcttttatcgtgatggagtaa
 gcgaagggaattctatcaagttttactctatgagttggatgcaattcgaaaggcttgtg
 catcgcttgaaccgaattatcagccaccgggtgacattcatagttgtacagaagcgtcacc
 acactcgtttgttctgtaataatcaccgagacaaaaacagtagtaccgaagcggaata
 tcttaccaggtactgtagttgacactaaaatattgtcatccaactgaattcgacttctacc
 tttgtagccatgcggtattcagggacaagcaggcctgcacattaccatgttctttggg
 acgagaacaatttcacagcagatggtattcaatctctgactaacaatctctgttatacct
 atgcgcggtgactcggtcggtctctatagtttctccagcgtattatgctcatcttgcag
 catttcgagcagtttctacttggaacctgagataatgcaagacaacggatcaccgggta
 aaaagaacacgaaaacaactctcgagacgtaggtgtgaagcctttaccagccttga
 aggagaatgtgaagagagtaattgttctactgctaaaaatccaaacattccttaatacagtt
 ttaataagtagtttgggtgttctgttagttcggttttagatttaccatgtttttctt
 atgtaaattttgcggtttggttaagccttttaggaattagtgattagggtttttctaa
 agttgtacttttagctgatgataacggttgatgcagtgactttgttaaaacctcctctcta
 cagtagtgtttacgtcgttctc

>G1146 Amino Acid Sequence (domain in AA coordinates: 886-896)
 MPIRQMKDSSETHLVIKTQPLKHHNPKTVQNGKIPPPSPSPVTVTPATVTQSQASSPSP
 PSKNRSRRNRNGGRKSDQGDVCMRPSRRPKPPPPSQTSSAVSVATAGEIVAVNHQMOM
 GVRKNSNFAPRPGFGLTKCIVKANHFLADLPTKDLNQYDVTITPEVSSKSVNRAIAE
 LVRLYKESDLGRRLLPAYDGRKSLYTAGELPFTWKEFSVKIVDEDDGIINGPKRERSYKVA
 IKFVARANMHLGEFLAGRADCPQEAQVILDIVLRELSVKRFCPVGRSFFSPDIKTPQR
 LGEGLBSWCGFYQSIRPTQMGSLNIDMASAAFIEPLPVIEFVAQLLGKDVLSKPLSDSD
 RVKIKKGLRGVKVEVTHRANVRKRYRVLGTTQPTRELMFPVDENCTMKSVIEYFQEMYG
 FTIQHTHLPCLQVGNQKKASYLPMACKIVEGQRYTKRLNEKQITALLKVTQRAEGQRN
 DILRTVQHNAYDQDPYAKEFGMNISEKLASVEARILPAPWLKYHENGKEKDCLPQVGQWN
 MMNKKMINGMTVSRWACVNFVRSVQENVARGFCNELGQMCEVSGMEFNPEPVIPIYSARP
 DQVEKALKHVYHTSMNKTGKLELELLAILPDNNGSLYGLDKRICETELGLISQCCLTKH
 VFKISKQYLADVSLKINVKMGRNTVLVDAISCRIPLVSDIPTIIFGADVTHPENGEES
 PSIAAVVASQDWPEVTKYAGLVCAQAHQELIQDLYKTWQDPVRGTVSGGMIRDLLISFR
 KATGQKPLRIIFYRDGVSEGFYQVLLYELDAIRKACASLEPNYQPPVTFIVVQKRHHTR
 LFANNHRDKNSTDRSGNILPGTVVDTKICHPTEDFYLCSHAGIQGTSRPAHYHVLWDEN
 NFTADGIQSLTNLCYTYARCTRSVSIVPPAYYAHLAAFRRFYLEPEIMQDNGSPGKKN
 TKTTTVGDVGKPLPALKENVKRMFYC*

>G1267 (152..967)
 AAGTAGAGAATAATAATCACATCAAGATTGTTTATAACCCCTCCCNTAATCACCTTCTTA
 NTNACCACCTCTCCGGCTCTCAACAGAACAAACAAAAACAGCTTCCGTTGTCTCTG
 TTCGGCGAAATCGGACGGTCGAGATCAATCATGCATCGTAGAGCAGCAATCAAGAATC
 GGATGACGAAGAAGATGAGACTTACAACGACGTCGTTCTCTGAATCTCCTTCTTCTGTGA
 AGACTCAAAGATCTCAAAACCACTCCAAAGAAAAGGAGGAACGTAGAGAAGAGAGTTGT
 CTCAGTTCCGATAGGTGACGTGGAAGGATCTAAGAGCAGAGGCGAAGTATATCCACCGTC
 CGATTCTATGGGCTGGAGAAAGTACGGACAAAAACCGATCAAAGGCTCGCCTTATCCCAG
 GGGATATTACAGATGTAGTAGCTCAAAGGATGTCCGGCGAGGAAGCAGGTGGAGAGAAG
 CCGTGTGGACCTTCTAAGCTTATGATTACTTACGCCTGCGACCACAATCACCTTTCCC
 TTCTCTCCGCTAACCACCAATCCCACCACCGCTCCTCCGTCGTCTCAAAACCGCAAA
 GAAAGAGGAAGAATACGAAGAGGAGGAAGAAGAACTAACCGTCACCGCGCGAGAGGAACC
 ACCGGCGGACTTGATCTAAGCCACGTAGACTCACCGTTGCTATTAGGCGGCTGCTACAG
 CGAAATCGGAGAGTTCCGGTGGTTCTACGACGCGTCGATCTCATCATCTGGTTCTTC
 GAATTTCTCCGACGTAACCTTAGAGAGAGGTTTTTTCAGTAGGCCAAGAGGAAGATGAGTC
 TTTGTTCCGTGATCTCGGTGATTTACCTGATTGCGCCTCCGTGTTCCGCCGTGGGACTGT

TGCGACGGAGGAGCAACATCGAAGATGTGATTTTGGCGCCATTCTTTCTGTGATAGTTC
TAGATGAGTTTGTGTGTGTAGCCAAACCAAAGAAAAACACAATTTTTTTATTTTCC
ACTGTAAAGGTGTATCAATGGTGGATTCAATTTTTTAAAAA

>G1267 Amino Acid Sequence (domain in AA coordinates: 70-127)
MHRRAAIQESDDEEDETYNDVVPESPSSCEDSKISKPTPKKRRNVEKRVVSVPIADVEGS
KSRGEVYPPSDSWAWRKYGQKPIKGSYPYRGYYRCSSSKGCPARKQVERSVDPSKLMIT
YACDHNHPPSSSANTKSHHRSSVVLKTAKKEEYEEEEELTVTAEEPPAGLDLSDVD
SPLLLGGCYSEIGEFWFYDASISSSSGSSNFDVTLERGFVSGQEEDESLFGDLGLDLPD
CASVFRRTGVATEEQHRRCDFGAIPFCDSSR*

>G1269 (88..951)
AACAAATCTCTCTCTCTTTATTCTTCTTCTTCAGCTTCAGATTTAGATCTTAAATCTTC
AAGTCTTCTTCTTCTTCTTCTGCAACCATGGCTATGCAGGAACGTTGTGAGAGTTTATGT
TCTGATGAACCTTATATCTTCTCAGATGCCTTTTACCTCAAGACAAGAAAGCCTTATACC
ATCACTAAACAAAGAGAGAAATGGACAGAAGCAGAGCATGAGAAGTTTGTAGAAGCATTG
AAACTCTATGGCAGAGCTTGGAGACGAATCGAAGAACATGTTGGAACAAAACTGCAGTT
CAGATTGGAAGCCATGCGCAGAAGTTCTTTACTAAGGTTGCTCGCGATTTTGGTGTAGC
TCTGAGTCCATTGAGATCCCGCTCCAGGCCAAAGAGAAAGCCGATGCATCCTTACCCT
AGAAAGCTTGTGATTCCTGATGCAAAAGAGATGGTATACGCTGAACCTAACCGATCCAAG
CTGATTGAGGATGAAGATAACCGATCTCCAACATCGGTTTTATCAGCTCATGGCTCAGAT
GGATTAGGTTCCATTGGTTCAAATTCACCTAATCTTCTTTCAGCTGAGTTATCATCTCAC
ACAGAGGAATCATTTGCTCTCTGAAGCAGAGACCAACAGAGCCTTAAGCTCTTTGGAAAA
ACTTTTGTAGTTGGTGATTACAATCTTCAATGAGTTGTGATGATTCTGAAGATGGCAAG
AAGAAGCTATACTCAGAAACACAGTCTCTTCAATGTTCTTCTTCTACTTCAGAAAACGCT
GAAACAGAAGTGGTAGTGTGCGAGTTCAAAGAAGTGAGAGATCAGCTTTCTCTCAGTTA
AAATCGTCGGTGACTGAGATGAACAACATGAGAGGGTTTCATGCCTTACAAAAGAGAGTA
AAGGTGGAAGAAACATTGACAATGTAAATATCATATCCTTTGTGGTGAAGTGTTCGT
TTGTGTCAAGTCAGTTGTGTAACCTTTTGTGATCTCAACATCAGATTATGTGTATAATGT
CAGAGTATTAGGGAAAGTTTTTTTGGATTAGATTGTAAGATCACTCCAAAGTTTCGTGT
CTTTCCATATAACCGATTAGAAATGAGATCCTTGTACTTAAACATTTTTATTTGATCAA
TCAAATCTTCTTGATGAAAAA

>G1269 Amino Acid Sequence (domain in AA coordinates: 27-83)
MAMQERCELSDELISSSDAFYLKTRKPYTITKQREKWTEAEHEKFVEALKLYGRAWR
IEEHVGTKTAVQIRSHAQKFFTKVARDFGVSSSEIEIPPRPKRKMHPYPKRLVIPDAK
EMVYAEITGSKLIQDENDNSPTSVLSAHGSDGLSIGSNSPNSSSAELSSHTESLSLEA
ETKQSLKLFKTFVVDYNSMSDCDDSEDGKKLYSETQSLQCSSSTSENAETEVVVSEF
KRSERSAFSQLKSSVTEMNMRGFMPYKKRVKVEENIDNVKLSYPLW*

>G1452 (175..1296)
ATTTATTAAGCATCAATGAGAGAACTTCAGAGCTGGGTTTGAGTTCTGTCCAATAATACA
TAACCACGTTATCATTTTTTGTCTTTACTATCTCATTACACTCTTCTGTTATTCGCCCAA
TTCTTACAGTCATTACTCTCTATAGGGCTCGAGCGGCCCGCCGGGAGGTTTCTATGCAG
ATGGTTACACTTCCCGCTCCATTGCCAGATTGGGTTTCGGTGTAAAGTCGCAATTAGTA
CTCACTATAGGGCTCGAGCGGCCCGCCGGGAGGTAAAGATCAAACAATGTCTAAAGAA
GCTGAGATGTCGATCGCGGTGTCGGCTTTGTTCCCTGGTTTATAGATTCTCTCTACTGAT
GTTGAACCTTATCTCGTACTATCTTCGTCGTAATAATCGATGGTGATGAGAACTCTGTTGCT
GTGATTGCTGAGGTCGAGATTTACAAGTTTCGAGCCGTGGGACTTGCCAGAGGAATCGAAA
CTGAAATCGGAGAACGAGTGGTTTTACTTCTGCGCGAGGGGAGGAAGTACCCGCACGGG
TCACAAAGCCGCGAGCCACACAGCTAGGATATTGGAAGCGACCGGTAAAGAGCGGAGT
GTTAAATCCGGAACCAAGTTGTTGGAACCAAGAGAACGCTTGTATTTTCATATCGGTCCG
GCTCCTCGTGGCGAGAGAACGGAGTGGATTATGCATGAATACTGCATCCATGGAGCCCCA
CAGGATGCATTAGTGGTGTGCCGGTTAAGAAAAATGCTGATTTTCGGGCTAGTTTCGACC
CAAAAAATTGAGGATGGTGTGTGCAAGACGATGGCTACGTTGGCCAAAGAGGTGGTTTG
GACAAGGAGGACAAATCCTACTATGAATCTGAGCATCAGATACCAATGGTGACATCGCA
GAATCATCAAATGTTGTTGAGGATCAGGCCGATACCGATGATGATTGTTACGCCGAGATT
TGAACGATGATATAATAAGCTCGACGAAGAAGCGTTGAAAGCTAGCCAAGCGTTTCGA
CCAACCTAATCCAATCATCAAGAAACAATATCAAGCGAGTCATCGAGTAAGAGGTCAAAA
TGTGGTATAAAAAAGAAATCAACGGAACAATGAATTGTTACGCTTTGTTTCAGGATCAAG
AACGTTGCCGAACCGACTCCAGCTGGAGATTCGGAACCCGTTCAAATCAAGAAAGAT

GATAGCCAGAGATTGATGAAGAATGTTCTGGCCACTACTGTTTTCTTGGCTATCTTATTT
TCTTTCTTTTGGACTGTATTAATAGCTAGGAACTAAAGCTAGTTACGACATACATATTAT
TTATACATAAAATAAATATAGTATTTTGTCTATGGCAAAAAAAAAAAAAAAAAA

>G1452 Amino Acid Sequence (domain in AA coordinates: 30-177)
MQMVHTSRISIAQIGFGVKSQVLVTIGLERPPGQVKDQMSKEAEMSIASVAFPGFRFSP
TDVELISYYLRRKIDGDENSVAIVAEVEIYKFEPWDLPEESKLKSENEWFYFCARGRKYP
HGSQSRRTQLGYWKATGKERSVKSGNQVVGTKRTLVPFHIGRAPRGERTEWIMHEYCIHG
APQDALVVCRLRNADFRASSTQKIEDGVVQDDGYVGQRGGLDKEDKSYESEHQIPNGD
IAESSNVVEDQADTDDDCYAEILNDDIIKLDEEALKASQAFRPTNPHTQETISSESSSKR
SKCGIKKESTETMNCYALFRIKNVAGTDSWRFPNPFKIKKDDSQRLMKNVLATTVFLAI
LFSFFWTVLIARN*

>G1494 (114..1406)

TCGACAGAGTTGTGTTGGGCGTGGAACCTTGGACTAGTTCCACATATCAGGTTATATAGAT
CTTCTCTTTCAACTTCTGATTTCGTCCAGAAGCTTTCTTAATCTGAGATCTGACATGGAAC
ACCAAGGTTGGAGTTTGGAGGAGAATTATAGTTTGTCCACTAATAGAAGATCTATCAGGC
ACAAGATGAAGTGTGGAGTTATTATGGCGAGATGGACAAGTGGTTCTGCAGAGCCAAA
CTCATAGAGAACAAACCCAAACCAAGATCATCATGAAGAAGCCCTAAGATCCA
GCACCTTTCTTGAAGATCAAGAACTGTCTCTTGGATCCAATACCTCCAGATGAAGACC
CATTGCAACCCGACGACTTCTCTCCCACTTCTTCTCAACCATGGATCCCCTCCAGAGAC
CAACCTCAGAGACGGTTAAGCCTAAGTCCAGTCCCTGAACCTCCTCAAGTCATGGTTAAGC
CTAAGGCCTGTCTGACCTCTCTCTCAAGTCATGCCTCCTCCAAAATTTAGGTTAACAA
ATTCATCATCGGGGATTAGGGAACAGAAATGGAACGACTCTCGGTAACGACCGTTGGAC
CTAGCCATTGCGGAAGCAACCCATCACAGAACGATCTCGATGTCTCAATGAGTCATGATC
GAAGCAAAAACATAGAAGAAAAGCTTAATCCGAACGCAAGTTCTCATCAGGTGGCTCCT
CTGGTTGCAGCTTTGGCAAAGATATCAAAGAAATGGCTAGTGGAAAGATGCATCACAAACG
ACCGTAAGAGAAAACGTATAAATCACACTGACGAATCTGTATCTCTATCAGATGCAATCG
GTAACAAGTCGAACCAACGATCAGGATCAAACCGAAGGAGTCGAGCAGCTGAAGTTCATA
ATCTCTCCGAAAGGAGGAGGAGAGATAGGATCAATGAGAGAATGAAGGCTTTGCAAGAAC
TAATACCTCACTGCAGTAAACTGATAAAGCTTCGATTTTAGACGAAGCCATAGATTATT
TGAAATCACTTCAGTTACAGTTCAAGTGATGTGGATGGGGAGTGGAAATGGCGGCGGCGG
CGGCTTCGGCTCCGATGATGTTCCCGGAGTTCAACCTCAGCAGTTCATACGTCAGATAC
AGAGCCCGGTACAGTTACCTCGATTTCCGGTTATGGATCAGTCTGCAATTCAGAACAAATC
CCGGTTTAGTTTGGCAAACCCGGTACAAAACAGATCATCTCCGACCGGTTTGCTAGAT
ACATCGGTGGGTTCCCAACATGCAGGCCGCGACTCAGATGCAGCCGATGGAGATGTTGA
GATTTAGTTACCGGCGGGACAGCAAAGTCAACAACCGTCGTCTGTGCCGACGAAGACCA
CCGACGGTTCTCGTTTGACCCTAGGTTGGTGAGCCACTTTGC

>G1494 Amino Acid Sequence (domain in aa coordinates: 261-311)
MEHQGWSFEENYSLSNRRSIRPQDELVELLWRDQVVLQSQTTHREQTQTQKQDHHEAL
RSSTFLEDQETVSWIQYPPDEDPFEPDDFSSHFSTMDPLQRPSTETVKPKSSPEPPQVM
VKPKACPDPPPQVMPKFRLLTNSSSGIRETEMEQYSVTTVGPSHCGSNPSQNDLDVSMS
HDRSKNIEEKLNPNASSSSGSGSGSFGKDIKEMASGRCTTDRKRKRINHTDESVSLSLSD
AIGNKSNQSRSGSNRRSRAAEVHNLSERRRRDRINERMKALQELI PHCSKTDKASILDEAI
DYLKSLQLQLQVMWMSGMAAAAASAPMMFPGVQPPQFIRQIQSPVQLPRFPVMDQSAIQ
NNPGLVCQNPVQNQIISDRFARYIGGFPHMQAATQMOPMEMLRFSPPAGQQSQQPSSVPT
KTTDGSRLDH*

>G1548 (1..2511)

ATGGCAATGTCTTGCAAGGATGGTAAGTTGGGATGTTTGGATAATGGGAAGTATGTGAGG
TATACACCTGAACAAGTTGAAGCACTTGAGAGGCTTTATCATGACTGTCTAAACCGAGT
TCTATTGCGCGTCAGCAGTTGATCAGAGAGTGTCTATTCTCTTAACATTGAGCCTAAA
CAGATCAAAGTGTGGTTTTCAGAACCGAAGATGTAGAGAGAAACAAAGGAAAGAGGCTTCA
CGGCTTCAAGCTGTGAATCGGAAGTTGACGGCAATGAACAAGCTCTTGATGGAGGAGAAT
GACAGGTTGCAGAAGCAAGTGTACAGCTGGTCCATGAAAACAGCTACTTCCGTCAACAT
ACTCCAAATCCTTCACTCCAGCTAAAGACACAAGCTGTGAATCGGTGGTGACGAGTGGT
CAGCACCATTGGCATCTCAAAATCCTCAGAGAGATGCTAGTCTGCAGGACTTTTGTCC
ATTGCAGAAGAACTTTAGCAGAGTTTCTTTCAAAGGCAACTGGAACCGCTGTTGAGTGG
GTTCAGATGCCTGGAATGAAGCCTGGTCCGATTCATTGGAATCATCGCTATTTCTCAT
GTTGCACTGGTGTGGCAGCACGCGCCTGTGGCCTAGTGGGTCTTGAGCCTACAAGGGTT

GCAGAGATTGTCAAGGATCGTCCTTCGTGGTTCCGCGAATGTCGAGCTGTTGAAGTTATG
AACGTGTTGCCAACTGCCAATGGTGGAAACCGTTGAGCTGCTTTATATGCAGCTCTATGCA
CCAACCTACATTGGCCCCACCAACGCGATTCTTGGCTGTTACGTTACACCTCTGTTTTAGAA
GATGGCAGCCTTGTGGTGTGCGAGAGATCTCTTAAGAGCACTCAAAATGGTCCTAGTATG
CCACTGGTTCAGAATTTTGTGAGAGCAGAGATGCTTTCCAGTGGGTACTTGATACGGCCT
TGTGATGGTGGTGGCTCAATCATACACATAGTGGATCATATGGATTGAGGCTTGTAGC
GTGCCTGAGGTCTTGCGCCCGCTCTATGAGTCACCCAAAGTACTTGACAGAAAGACAACA
ATGGCGGCACTGCGTCAGCTCAAGCAAATAGCTCAGGAGGTACTCAGACTAATAGTAGT
GTTAATGGGTGGGGACGGCGTCCTGCTGCCTTAAGAGCTCTCAGCCAGAGGCTAAGCAGA
GGCTTCAATGAAGCTGTAAATGGTTTCACTGATGAAGGATGGTCAGTGATAGGAGATAGC
ATGGATGATGTCACAATCACTGTAAACTCTTCTCCAGACAAGCTAATGGGTCTAAATCTT
ACATTTGCCAATGGCTTTGCTCCTGTAAGCAATGTTGTTTTATGCGCAAAAGCATCAATG
CTTTTACAGAAATGTTCTCCGGCGATCCTGCTTCGGTTTTCTGAGGGAGCATAGGTGAGAA
TGGGCTGACAAACATTGATGCGTATCTAGCAGCAGCAGTTAAAGTAGGGCCTTGTAGT
GCCCCAGTTGGAGGATTGAGGGGCGAGTTATACTTCCACTTGCTCATACTATTGAGCAT
GAAGAGTTTATGGAAGTCATCAAATTGGAAGGTCTTGGTCATTCCCCTGAAGATGCAATC
GTTCCAAGAGATATCTTCCCTTCTCAACTTTGTAGCGGAATGGATGAAAATGCTGTAGGA
ACCTGTGCGGAACCTTATATTGCTCCAATCGATGCTTCGTTTGGCGGATGATGCACCTCTG
CTTCCCTTCTGGTTTTCTGATTTATCCCTCTTGATTCCGCAAAGGAAGTATCTAGCCCAAAC
CGAACCTTGGATCTTGCTTCGGCACTGGAAATTGGTTCAGCTGGAACAAAAGCCTCAACT
GATCAATCAGGAACTCCACATGTGCAAGATCTGTGATGACAATAGCATTTGAGTTTGGT
ATCGAGAGCCATATGCAAGAACATGTAGCATCCATGGCTAGGCAGTATGTTTCGAGGTATC
ATATCATCGGTGCAGAGAGTAGCATTGGCTCTTCTCCTTCTCATATCAGCTCAAGATT
GGTCTACGCACTCCTTTGGGTACTCCTGAAGCCCAAACACTTGCTCGTTGGATTGCCAG
AGTTACAGGGGCTACATGGGTGTTGAGCTACTTAAATCAAACAGTGACGGCAATGAATCT
ATTCTTAAGAATCTTTGGCATCACACTGATGCTATAATCTGCTGCTCAATGAAGGCCTTG
CCCGTCTTACATTTGCAAACAGGCGGGACTTGACATGCTGGAGACTACATTAGTTGCT
CTTCAAGACATCTCTTAGAGAAGATATTTGATGACAATGGAAGAAAGACTCTTTGCTCT
GAGTTCCACAGATCATGCAACAGGGCTTCGCGTGCCCTCAAGCGGGATATGTCTCTCA
AGCATGGGGAGACAGTTTCGTATGAGAGAGCAGTTGCTTGGAAAGTACTCAATGAAGAA
GAAAATGCTCATTGCATCTGCTTTGTGTTTCATCAATTGGTCTTTGTGTGA

>G1548 Amino Acid Sequence (domain in AA coordinates: 17-77)

MAMSKDKGLGLDNGKYVRYTPQEVALERLYHDCPKPSSIRRQQLIRECPILSNIEPK
QIKVWFQNRRCREKQRKEASRLQAVNRKLTAMNKLMBENDRLQKQVSQLVHENSYPFROH
TPNPSLPKADTSCSVVTSQGHLASQNPQRDASPAGLLSIAEETLAEFLSKATGTAVEW
VQMPGMKPGPDSIGIIAISHGCTGVAARACGLVGLLEPTRVAEIVKDRPSWFRECRAVEVM
NVLPTANGGTVELLYMQLYAPTTLAPPRDFWLLRYTSVLEDGSLVVCERSLKSTQNGPSM
PLVQNFVRAEMLSSGYLIRPCDGGGSIHIVDHMDLEACSVPEVLRPLYESPKVLAQKTT
MAALRQLKQIAQEVTTQNSSVNGWGRRPAALRALSQLSRGFNEAVNGFTDEGWSVIGDS
MDDVTITVNSSPDKMLMNLTFANGFAPVSNVVLCAKASMLLQNVPPAILLRFLREHRSE
WADNNIDAYLAAAVKVGPC SARVGGFGQVILPLAHTIEHEEFMEVILEGLGHSPEDAI
VPRDIFLLQLCSGMDENAVGTCAELIFAPIDASFADDAPLLPSGFRIIPLDSAKEVSSPN
RTLDLASALEIGSAGTKASTDQSGNSTCARSVMTIAFEFGIESHMQEHVASMARQYVRGI
ISSVQRVALALSPSHISSQVGLRTPGTPEAQTARWICQSYRGMGVVELLSNSDGNES
ILKNLWHHTDAIICCSMKALPVFTFANQAGLDMLETTTLVALQDISLEKIFDDNGRKTLCs
EFPQIMQQGFACLOGGICLSSMRPVS YERAVAWKVLNEENAHICFPVINWSFV*

>G1574 (1..1962)

ATGGATGATACAATGGACATGAGTTCAGGTAGTGATGAAGAAGTACAAGAAGAGAAGACC
ACTGTTAACGAGAGGGTCACTATCAGGCTGCATTACAAGATCTGAAGCAACCCAAGACC
GAAAAGGATCTACCTCCTGGTGTCTTACAGTTCCTCTTATGAGGCATCAGAAAATTGCA
TTGAACTGGATGCGTAAGAAAGAAAAAGAACAGGCACTGTTTGGGAGGGATATTAGCA
GATGATCAGGCACTTGGTAAACGATCTCGACGATCTCTCTATCCTGTTACAAAAGTTG
AAGTACAATCAAAGCAGAGAAAGCGAAAAGGTCAAAACTCTGGTGGTACATTGATTGTT
TGTCCAGCAAGTGTGTAAACAATGGGCAAGAGAAGTTAAAGAGAAGGTTTCTGATGAA
CACAACTCTCTGTTTTAGTCCACCATGGATCTCACAGAACCAAGATCCACAGAAATA
GCAATATATGATGTGGTCATGACAACTTACGCCATTGTTACAAATGAAGTTCCACAAAAC
CCTATGCTGAATCGTTATGATAGTATGAGAGGCAGAGAAAGCCTTGACGGATCGAGTTTG

ATTACGCTCACGTTGGTGCCTAGGAAGAGTTAGGTGGTTGAGAGTAGTATTAGATGAA
GCTCATACAATTAAAAACCATAGAACCTTAATTGCAAAAGCTTGTTTTAGCCTTAGAGCC
AAAAGGAGATGGTGTGTTGACTGGAACGCCGATAAAGAACAAAGTAGACGATCTTTATAGC
TATTTTCAGATTTCTTAGATATCATCCATATGCCATGTGCAATTCATTTACCAAAGAATC
AAAGCTCCAATTGATAAAAGCCTCTTCATGGTTACAAGAAGCTTCAAGCTATTCTAAGG
GGTATAATGTTGCGCCGCACCAAAGAATGGTCTTTCTACAGGAAGCTTGAATTGAATTCA
CGTTGGAAGTTTGAGGAATATGCTGCTGATGGGACTTTGTCATGAACACATGGCTTATCTT
TTGGTGATGCTTTTGCGACTACGCCAAGCTTGTAAACCATCCACAACCTTGTTAACGGATAT
AGTCACTCAGATACTACAAGAAAAATGTCAGATGGAGTTCGAGTAGCCCCCTAGAGAGAAT
CTAATCATGTTCTCGATCTCTTGAAATTATCCTCAACCACCTGCTCTGTTTGTAGTGAT
CCACCAAAGACCCTGTTGTTACTTTGTGTGGCCATGTGTTTTGTTATGAGTGTGTGTCT
GTAAACATTAAACGGGGATAACAATACGTGCCCTGCCTTAATTGCCACAGCCAGCTTAAA
CATGATGTTGTTTTCACTGAATCTGCAGTTAGAAGTTGCATCAACGATTATGATGATCCT
GAAGATAAAAATGCTTTAGTTGCATCAAGCGAGTTTATTTTCATCGAAAAATCCGAGCTGT
GATAGAGATTCTTCAGTCGCTTGACAGCAAGGCAGTCCAGACACTCCACCAATAAAGAC
AATAGTATCAGTGGACTGAATCTCATTCTTACGTTTCTCAAAGACAAATGTAATGATTAT
GAAACAGGTGCGATGTTGATGTCTCTTAAAGCTGGAACCTTGGATTGAATATGGTAGCT
GCAAGTCATGTCATTCTACTGGACCTATGGTGAATCCAACAACAGAGGATCAAGCTATT
GATCGAGCTCATCGTATCGGACAACTCGAGCTGTTACGGTCACTCGTATTGCCATCAA
AATACCGTTGAGGAACGAATTTTGACTCTTCATGAACGTAAAGGAACATTGTTGCATCT
GCATTGGGTGAAAAAACTGGCAAAAGTTCTGCGATTCAACTAACACTAGAAGATCTCGA
ATATCTGTTTTTTGGTGTGTAGAATATCCAGAGTTTTTTATTGATAAGAGGAATAAAACC
TTTAGCTATTTAATAAGTCACAAGTGTGAATGAATGAATAA

>G1574 Amino Acid Sequence (domain in AA coordinates: 28-350)
MDDTMDMSSGSDEEVQEEKTVNERVIYQAALQDLKQPKTEKDLPPGVLTVPPLMRHQKIA
LNWMRKKEKRSRHLGLGILADDQGLKTIISTISLILLQKLKSQSKQRKRKGQNSGGTLIV
CPASVVQWAREVKEKVSDEHKLSVLVHHGSHRTKDPTEIAIYDVVMTTYAIVTNEVPQN
PMLNRYDSMRGRESLDGSSLIQPHVGALGRVRLRVVLDEAHTIKNHRTLIAKACFSIRA
KRRWCLTGTPIKNKVDLYSYFRFLRYHPYAMCNSFHQRIKAPIDKPLHGYKKLQAILR
GIMLRRTKEWSFYRKLELNSRWKFEEYAADGTLHEHMAILLVMLRLRLQACNHPQLVNGY
SHSDTTRKMSDGVVRVAPRENLMFLDLLKLSSTTCSVCSDDPKDPVVTLCGHVFCYECVS
VNINGDNNTCPALNCHSQLKHDVVFTESAVRSCINDYDDPEDKNALVASRRVYFIENPSC
DRDSSVACRARQSRHSTNKDINSISGLNLIFTFLKDKCNDYETGAMLSLKAGNLGLNMVA
ASHVILLDLWNPPTTEDQAIIDRAHRIGQTRAVTVTRIAIKNTVEERILTLHERKRNIVAS
ALGEKNWQKFCDSNTNRRSRISVFWCVEYPRVFIIDKRNKTFSYLISHKCECNE*

>G1586 (1..807)
ATGAATCAAGAAGGTGCTTCACATAGCCCATCCTCCACTTCCACCGAACAGTCCGGGCA
CGTTGGTCACCTAAACCGGAGCAAATCTTGATACTCGAATCCATCTTCAACAGTGGTACT
GTTAACCACCAAAGATGAAACGGTGAGGATAAGAAAGATGCTTGAGAAATTCGGTGCT
GTGGGAGACGCAACGCTCTTCTACTGGTTTCAAACCGACGGTCAAGATCTCGCCGGAGA
CACCGGCAGCTTTTAGCAGCCACCACCGCAGCCGCCACCTCCATAGGAGCTGAAGACCAC
CAGCACATGACGGCCATGAGCATGCATCAATATCCTTGACAGCAACAACGAGATTGATTTG
GGTTTGGAAGTTGTAGCAACTTATCAGCTAATTACTTCTTAATGGATCGTCGTCATCT
CAAATCCCTTCTCTTTCTCCTCGGCTCTCTCTTCTCAAGTGGTGGGTGTGAGAACAAAT
GGTATGGAGAATCTCTTCAAATGTATGGCCATGAATCTGATCATAATCATCAGCAGCAG
CATCATAGCTCAAATGCTGCATCAGTTTTAAACCCATCTGATCAAACTCCAATCCCAA
TACGAACAAGAAGGGTTTATGACGGTGTATATAACGGAGTTCCTATGGAAGTAACAAAA
GGAGCAATAGACATGAAACAATGTTCCGGTGATGATTCGGTGTACTTCATTCCTCTGGT
CTTCCTCTTCCCACTGATGAGTTTGGTTCTTGTATGCATTCTTTACAACATGGACAACT
TATTCCTGGTACCGAGACAGACATGA

>G1586 Amino Acid Sequence (domain in AA coordinates: 21-81)
MNQEGASHSPSSTSTEPVRARWSPKPEQILILESIFNSGTVPNPKDETVRIRKMLEKFGA
VG DANVFYWFQNRSSRRRRHRQLLAATTAATSIGAEDHQHMTAMSMHQYPCSNNEIDL
GFGSCSNLSANYFLNGSSSSQIPSFLLGLSSSSGGCENNGMENLFKMYGHESDHNHQQQ
HHSSNAASVLNPSDQNSNSQYEQEGFMTVFINGVPMVETKGAIDMKTMFGDDSVLLHSSG
LPLPTDEFGFLMHSLOHQGTYFLVPRQT*

>G1786 (1..1170)

ATGATCGTGACGGTGGGGGAGCATCCGAGGACGGTGAAGGTGGAGGGGTGGTTCTGAAG
AAAGGGCCATGGACGGTGGCCGAGGACGAGACACTGGCGGCTTACGTACGGGAATACGGT
GAAGGGAACCTGGAATTCGTTCAGAAGAAGACATGGCTGGCTAGGTGTGGCAAGAGCTGC
CGCCTCCGCTGGGCTAACCCTTACGACCTAATCTCAGGAAAGGCTCCTTACCCCCGAG
GAAGAACGTCTCATACAACTCCACTCTCAGCTAGGCAACAAATGGGCTCGCATGGCT
GCTCAGTTACCAGGCAGAACAGATAACGAGATCAAGAACTACTGGAACACGAGGTTGAAA
CGCTTCCAACGCCAAGGCCTCCCTCTCTACCTCCAGAATATCCCAAACAATCATCAA
CAACAAATGTATCCTCAACAGCCCTCCTCACCTCTCCCGTCCCAAACACCTGCTTCTTCC
TTTACCTTTCTCTCTCCAACCGCCTTCTCTGTGTCCCAAACGTTGTTATAACACTGCC
TTCTCTCCCAAGGCCTCATATATTTCTTCTCCAACCAATTCCTTGTCTCGTCTCCGACC
TTCTTTCACACCCATTCTCTCTTCTCTCTTATCAGTCTACCAATCCGGTTTACTCCATG
AAACATGAGCTCTCTTCAAACCAATTCCTACTCTGCCTCTTTAGGAGTCTATCAAGTA
AGCAAGTTCTCAGACAATGGGGATTGTAACCAAAACCTGAACACCGGTTTGCATACAAAT
ACCTGTCTCAGCTGTTAGAGGATCTTATGGAGGAGGCCGAGGCTCTAGCTGATAGCTTTCGT
GCTCCTAAGCGGAGACAAATCATGGCTGCGCTTGAGGACAACAACAACAACAACACTTT
TTCTCGGGAGGTTTCGGACATCGTGTTCCTTCCAACAGTCTATGTTCTTGCAAGGTTTA
ACACCAAAGGAAGATGAGTCTCTCCAGATGAACACAATGCAAGATGAGGACATAACAAAG
CTTCTTGACTGGGAAGTGAAGAAATCTCAAACGGGCAATCCTCTGTGATAACA
ACAGAGAACAACCTTGTCTTGACGATCACCAGTTCGCTTTTCTGTTTCCAGTTGATGAT
GACACCAACAACCTTGCCAGGGATCTGCTAG

>G1786 Amino Acid Sequence (domain in AA coordinates: TBD)

MIYVGGGASEDEGGGVVLKKGPTVAEDELAAAYVREYEGGNWNSVQKKTWLRCKGKSC
RLRWANHLRPNLRKGSFTPEEERLI IQLHSQLGNKWARMAQLPGRDNEIKNYWNTRLK
RFQRQGLPLYPPEYSQNNHQQMYPQQPSSPLPSQTPASSFTFLLQPPSLCPKRCYN TA
FSPKASYISSPTNFLVSSPTFLHTHSSLSYQSTNPVYSMKHELSSNQIPYSASLGVYQV
SKFSDNGDCNQNLNTGLHTNTCQLLEDLMEEAEALADSFAPKRRQIMAALEDNNNNNNF
FSGGFGHRVSSNSLCSLQGLTPKEDES LQNM TMQDEDITKLLDWGSESEBISNGQSSVIT
TENNLVLDHQAFLFPVDDDTNNLP GIC*

>G1792 (77..496)

AATCCATAGATCTCTTATTAATAACAGTGCTGACCAAGCTCTTACAAAGCAAACCAATC
TAGAACACCAAAGTTAATGGAGAGCTCAAACAGGAGCAGCAACAACCAATCACAAGATGA
CAAGCAAGCTCGTTTCCGGGGAGTTCGAAGAAGGCCTTGGGGAAAGTTTGCAGCAGAGAT
TCGAGACCCGTCGAGAAACGGTGCCCGTCTTTGGCTCGGGACATTTGAGACCGCTGAGGA
GGCAGCAAGGGCTTATGACCGAGCAGCCTTTAACCTTAGGGGTCATCTCGCTATACTCAA
CTTCCCTAATGATATTATCCACGTATGGACGACTACTCGCTTCGCCCTCCTTATGCTTC
TTCTTCTTCGTCGTCGTCATCGGGTTCAACTTCTACTAATGTGAGTCGACAAAACCAAAG
AGAAGTTTTTCAGTTTGAGTATTGGACGATAAGGTTCTTGAAGAACTTCTTGATTGAGA
AGAAAGGAAGAGATAATCAGGATTAGTTTGTGTTTGATATTTATGTGGCACTGTTGTGG
CTACCTACGTGCATATGTTGCATGTATAGGTCGCTTGATTAGTACTTTATAACATGCATG
CCACGACCATAAATTGTAAGAGAAGACGTACTTTGCGTTTTTCATGAAATATGAATGTTAG
ATGGTTTGAGTACAAAAA AAAAAAAAAAAAAAAAAA

>G1792 Amino Acid Sequence (domain in aa coordinates: 17-85)

MESSNRSSNNQSQDDKQARFRGVRRRPWGKFAAEIRDPSRNGARLWLGTFTAEAEARAY
DRAAFNLRLGHLAILNFPNEYYPMDYSLRPPYASSSSSSSSSGSTSTNVSQRNQREVFEF
EYLDKVL EELLDSEERKR*

>G1865 (48..899)

AAGAAGAGGACATGAAGCACAGAGATTCTGCAGACTGCAGGTGACCAATGGACACTTTAT
CAATAAAAACATACTACTACTCTCTTACACTTTCAATTTTCCAATACAAATCCCAATCT
TTAATCTCTCTTTCTTCTCATCTCTCTTTCTCTTCTCTCTTCATGGCTACAAGGATTC
CATTACAGAAATCACAATGGGAAGAACTTGAAACCAAGCTCTTGTTGTTCAAGTACTTAG
CTGCAAAATATGCCTGTTCCACCTCATCTTCTTCTCCTCATCAAAAGACCCCTTTCTCTTCT
CTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCAAGCTTCTTCTCTCCACTCTTTCTCCACACT
TTGGGTGGAATGTGTATGAGATGGGAATGGGAAGAAAGATAGATGCAGAGCCAGGAAGAT
GTAGAAGAACTGATGGCAAGAAATGGAGATGCTCTAAAGAAGCTTACCCTGACTCTAAGT
ACTGTGAGAGACATATGCATAGAGGCAAGAACCCTTCTTCTCAAGAAAGCCTCCTCCTA
CTCAATTCACTCCAAATCTCTTTCTCGACTCTTCTTCCAGAAGAAGAAGTGGATACA
TGGATGATTTCTTCTCCATAGAACCTTCCGGGTCAATCAAAGCTGCTCTGGCTCAGCAA

TGGAAGATAATGATGATGGCTCATGTAGAGGCATCAACAACGAGGAGAAGCAGCCGGATC
 GACATTGCTTCATCCTTGGTACTGACTTGAGGACACGTGAGAGGCCATTGATGTTAGAGG
 AGAAGCTGAAACAAAGAGATCATGATAATGAAGAAGAGCAAGGAAGCAAGAGGTTTATA
 GGTTCCTTGATGAATGGCCTTCTTCTAAATCTTCTGTTTCTACTTCACTCTTCATTGAT
 CATCTTTTGTCTTATAACCTTGTATTTCTTGTTAAGATGGTAATGCAAATT

>G1865 Amino Acid Sequence (domain in AA coordinates: 124-149)
 MDTLSEIKTYLLSYTFNFIPIPIFNLSFFFISLSLSLFMATRIPFTESQWEELENQALV
 FKYLAANMPVPPHLLFLIKRPFLFSSSSSSSSSSSSFFSPTLSPHFGWNVYEMGMGRKIDA
 EPGRCRRTDGKKWRCSKEAYPDSKYCERHMRGKKNRSSSRKPPPTQFTPNLFLDSSRRR
 RSGYMDDFFSIEPSGSIKSCSGSAMEDNDGSCRGINNEEKQPDHCFILGTDLRTREPR
 LMLEEKLKQRDHDNEEEQSGKRFYRFLDEWPSSKSSVSTSLFI*

>G1886 (43..909)
 AGGAAACATAAGTAATCGTTGCTTCGATCCTTTTGTACATGGATGGATCCTGAACAGGAA
 ATCTCAAACGAGACTTTGGAAACTATATTGGTAAGTTCAACAAAAGGAAGCAATAATAAC
 AATAAGAAAATGGAAGAAGAAATGAAGAAGAAAGTATCAAGAGGAGAATTAGGAGGTGAA
 GCTCAAAATTTGTCCAAGATGTGAATCTCCAAACACAAAGTTTTGTTACTACAACAACTAT
 AGTCTCTCACAACCTCGTTACTTCTGCAAATCTTGTCCGAGATATTGGACTAAAGCGCGT
 ACTCTTCGTAACGTTCCCGTCCGTGGTGGTTGCCGTGAAACAAACGATCCTCTTCCTCA
 GCTTTCTCCAAGAACAACAATAAGTCTATTAATTCCATACTGATCCACTTCAGAAC
 CCTTTAATTACGGGAATGCCACCATCATCTTTTGGTTATGATCACTCCATTGATCTCAAC
 CTCGCTTTCGCTACTCTCCAAAGCATCATTTATCTCTCAAGCTACTACGCCTTCTTTT
 GGGTTTGGAGGTGATCTTTCTATTATGGAACCTCAACGAATGATGTAGGGATCTTCGGA
 GGGCAAAACGGTACTTATAACAATAGTTTGTGTTATGGGTTTATGTCCGGAAATGGTAAT
 AATAATCAAAATGAAATCAAGATGGCTTCTACATTGGGGATGTCTTTGGAAGGAAACGAG
 AGAAAGCAAGAGAAATGTGAACAATAACAATAAATCAAGAGAAATCTAGCAAGGTGTTT
 TGGGGGTTTCCATGGCAGATGACCGGAGATTCCCGCGGAGTTGTACCGGAGATTGATCCC
 GGAAGGGAAGCTGGAATGGGATGGTTTCTCTTGAATAATGGTTTACTCAACACTCCT
 TTGGTCTAGCAGATCATTA

>G1886 Amino Acid Sequence (domain in aa coordinates: 17-59)
 MDPEQEIENETLETILVSSTKGSNNNNKKMEEEMKKKVSRLGELGGEAQNCPRCESPNTKF
 CYNNYSLSQPRYFCKSCRRYWTGKGLLRNVPVGGGCRNRKRSSSAFKNMNNKSNINFH
 TDPLQNPLITGMPSSFGYDHSIDLNLAFATLQKHLSSQATTPSFGFGDLISYGNSTN
 DVGIFGGQNGTYNNSLCYGFMSGNGNNNNQNEIKMASTLGMSEGNERNKQENVNNNNNNSE
 NPSKVFWGFPWQMTGDSAGVVEIDPGRESWNGMVSSWNNGLLNTPLV*

>G1933 (33..1418)
 AATTGAGATTAAAGTAATTTATCTTTTCAGAAAATGGCGGTTGAAGACGATGTATCTTTGA
 TAAGAACGACGACGTTAGTGGCACCAACAGACCCACGATTACAGTTCCCTCATAGACCTC
 CGGCGATCGAAACGGCGGCTATTTCTTTGGCGGTGGAGATGGGCTTAGTCTAAGCCAG
 GGCCACTTTCTTTTGTCTCTTCTTTGTTTGTGATAACTTCCCTGACGTCTTGACGCCGG
 ATAACCAACGGACGACGTCGTTTACTCAGCTTCTTAACGGAATATGTCGGTGTCTCCTG
 GTGGCGGAGGACGTTCAACGGCGGGGATGTTCCGCGGAGGAGGTCCGATGTTTACAATCC
 CTTCTGGTTTCAGCCCTTCTAGTCTTCTCACCTCGCCCATGTTCTTTCCCCCGCAGTCGT
 CAGCTCATACCGGCTTTATTCAACCACGGCAGCAGTCACAACCGCAACCACAACGACCAG
 ACACGTTTCTCACCATATGCCACCATCGACATCCGTCGCCGTCCATGGTCGTCAATCTT
 TAGACGTTTTCACAAGTAGATCAAAGAGCTCGAAACCATTATAATAATCCGGGGAATAACA
 ATAATAACCGGTCGTATAACGTTGTGAACGTTGATAAACCGGCGGATGACGGTTATAACT
 GGAGGAAGTACGGACAAAAGCCTATCAAAGGGTGTGAATATCCAAGGAGTTATTACAAAT
 GTACACATGTTAACTGTCCGTGAAGAAGAAAGTCGAACGGTCATCGGATGGACAGATCA
 CTCAGATCATTTACAAAGGTCAACATGATCACGAGAGGCCTCAGAATCGCCGTGGCGGTG
 GAGGCAGAGATTCCACTGAGGTTGGTGGTGCAGGGCAATGATGGAATCTAGTGATGATA
 GTGGTTATCGTAAGGATCATGATGATGATGATGATGATGAAGATGATGAAGATCTTC
 CGGCTTCAAAGATAAGAAGATAGACGGTGTGTGACGACTCACCGGACGGTGACCGGAGC
 CTAAGATTATCGTTTCAGACAAAAGTGAAGTCGATCTTCTCGACGATGGCTATAGGTGGC
 GTAAGTACGGACAAAAGTTGTCAAAGGAAATCCCATCCAAGGAGCTATTATAAATGTA
 CAACGCCAAAATTGTACGGTCCGTAAACATGTAGAGAGAGCTTCCACGGATGCTAAGGCTG
 TGATTACAACCTACGAAGGTAAACACAATCACGATGTCCCTGCCGCTAGAAACGGTACCG
 CGGCAGCAACCGCAGCTGCGGTGGGGCCGTCTGACCACCATCGTATGAGATCAATGTCCG

>G1933 Amino Acid Sequence (conserved domain in AA coordinates:205-263, 344-404)

>G2059 (58..1089)

>G2059 Amino Acid Sequence (conserved domain in AA coordinates:184-254)

>G2105 (42..1487)

46

TTGGTGTTCATATGGAGAGGTTGATAGAGAAGCAAGATTTTTGGCTTGAGAAGTTGATGA
 AGATTGTGGAAGACAAAGAACATCAAAGGATGCTGAGAGAAGAGGAATGGAGAAGGATTG
 AAGCGGAAAGGATCGATAAGGAACGTTTCGTTTTGGACAAAAGAGAGGGAGAGGATTGAAG
 CTCGGGATGTTGCGGTGATTAATGCCTTGCAGTACTTGACGGGAAGGGCATTGATAAGGC
 CGGATTCCTCGTCTCTACAGAGAGGATTAATGGGAATGGAAGCGATAAAATGATGGCTG
 ATAATGAATTTGCTGATGAAGGAATAAGGGCAAGATGGATAAAAAACAAATGAATAAGA
 AAAGGAAGGAGAAATGGTCAAGCCACGGAGGGAATCATCCAAGAACCAAAGAGAATATGA
 TGATATACAACAATCAAGAACTAAGATTAATGATTTTTGTGCGAGATGATGACCAATGCC
 ATCATGAAGGTTACTCACCTTCAAACCTCAAGAACGCAGGAACCTCCGAGCTGCAGCAATG
 CCATGGCAGCTAGTACAAAGTGCTTTCCATTGCTTGAAGGAGAAGGAGATCAGAACTTGT
 GGGAGGGTTATGGTTTGAAGCAAAGGAAAGAAAATAATCATCAGTAAGCTACATTTTTCA
 TTCTCAAATGAAGAATAAGAGAACTTAGAAACGAT

>G2105 Amino Acid Sequence (domain in AA coordinates: 100-153)

MEDHQNHPPQYQIEQPSSQFSSDLFGFNLVSAPDQHHRLHFTDHEISLLPRGIQGLTVAGN
 NSNTITTIQSGGCVGGFSGFTDGGGTGRWPRQETLMLLEVRSLDHKFKKANQKGPLWDE
 VSRIMSEEHGYTRSGKKCREKFNLYKYKTKKGGKSGRRQDGKNYRFFRQLEAIYGESK
 DSVSCYNNTQFIMTNALHSNFRASNIHNIVPHHQNPLMTNTNTQSLSISNNFNSSSDL
 DLTSSEGNETTKREGMHWKEKIKFVIGVHMERLIEKQDFWLEKLMKIVEDKEHQRLRE
 EEWRRIEAEERIDKERSFWTKERERIEARDVAVINALQYLTLGRALIRPDSSSPTERINGNG
 SDKMMADNEFADEGNKGMKMDKKQMNKKRKEKWSHGGNHPRTKENMMIYNNQETKINDFC
 RDDDQCHHEGYSPPNSKNAGTPSCSNMAASTKCFPLLEGEQDQNLWEGYGLKQORKENNH
 Q*

>G2117 (49..465)

ATACTTGTCAACAAAAATTTTCTTAAAGAACGCATAACTGTTTTTTCATGGCTGGTTCT
 GTCTATAACCTTCCAAGTCAAAACCCTAATCCACAGTCTTTATTCCAAATCTTGTGAT
 CGAGTACCACTTTCAAACCTTGCTGCCACGTCAGACGACTCTAGCCGGACTGCAGAAGAT
 AATGAGAGGAAGCGGAGAAGGAAGGTATCGAACCAGGAGTCAAGTCTCGGAGATCGCGTATG
 CGGAAACAGCGTCACATGGAAGAACTGTGGTCCATGCTTGTTCAACTCATCAATAAGAAC
 AAATCTCTAGTCGATGAGCTAAGCCAGCCAGGGAATGTTACGAGAAGGTTATAGAAGAG
 AACATGAAACTTCGAGAGGAAAACCTCAAGTTCGAGGAAGATGATTGGTGAGATCGGGCTT
 AATAGGTTTCTTAGCGTAGAGGCCGATCAGATCTGGACCTTCTAATCGTCTCGTAAGCTT
 GTTGGTTTTTTGTTGTTTATTTAAAG

>G2117 Amino Acid Sequence (conserved domain in AA coordinates: 46-106)

MAGSVYNLPSQNPQSLFQIFVDRVPLSNLPATSDSSRTAEDNERKRRRKVSNRESAR
 RSRMRKQRHMEELWSMLVQLINKNKSILVDELSQARECYEKVIEENMKLRENSKSRKMIG
 EIGLNRFLSVEADQIWT*

>G2124 (87..923)

GAACAGCAAAACCCTAGATTTCTGTTCAAGCTCAAGACCGTACAAAACCTTGGAACTCA
 TATATAAGATCTCGAGAATAGCATTATGAATATCGTCTCTTGGAAAGATGCAAACGACG
 AAGTTGCAGGCGGCGCTACGACAAGACGTGAAAGAGAAGTAAAGAGGATCAAGAAGAAA
 CCGAAGTCAGACCCACCACTAGTGGCAAACCGTAATTAAAAAGCAGCCTACATCGATCTCTT
 CTTCTTCTTCTTCTGATGAAATCCAAGGATCCGAGGATTGTTAGGGTTTCACGCGCCT
 TTGGAGGCAAAGACCGTCAAGCAAGTGTGTACGTTACGTGGACTACGTGACAGACGCG
 TGAGATTATCAGTCCCAACGCTATTACGCTCTACGATCTTCAAGAACGCTCGGTGTTG
 ACCAGCCTAGCAAAGCCGTTGACTGGTTGCTTGATGCAGCTAAAGAGGAGATCGACGAGC
 TACCTCCGTTACCTATCTCGCCGAAAATTTACGATCTTCAACCATCATCAGTCCTTCT
 TGAATCTTGGTCAACCGCCCGGTCAAGATCCGACCCAACTCGGGTTTAAATCAATGGAT
 GTGTACAAAGTCTACTACTACTAGCCGCGAAGAAAACGATAGAGAGAAAGGAGAAAACG
 ATGTCGTTTACACAAACAATCATCATGTTGGGTCTTATGGAACCTATCACAACCTGGAAC
 ATCATCATCATCATCAACATTTGAGTTTACAGGCAGATTATCATAGTCATCAACTAC
 ATAGTCTTGTCCCATTTCCATCACAATTTTGTATGTCCAATGACGACATCACCACAA
 CTACAACTATACAACTTTTGTTCATCATCATCGTCAGCTGGTTTCAGGGACTATGGAGA
 CATTAGATCCGAGGCAAATGTAGCAACAATGGTGGTAGAGACATTGATAATCGGATGTGCG
 TCGGTCCAATTCACCGAACTAATAGCACTACACCGGCTAACATGTGCGAGGCATCTAGGC
 TCGGAGCGTTGTACCAAGTAGGAAGTGATCACCATATGTGAAGTTAGATTATTGAAACG
 ATATAATTGTTGTTGATGTGTTTCAGAAATAAGGGGACAC

>G2124 Amino Acid Sequence (domain in AA coordinates: 75-132)

MNIVSWKDANDEVAGGATRRREREVKEDQEETEVRATSGKTVIKKQPTSISSSSSSSWMKS
KDPRIVRVSRAFGGKDRHSKVCTLRGLRDRRVRLSVPTAIQLYDLQERLGVDQPSKAVDW
LLDAAKEEIDELPPLPISPENFSIFNHHQSFLNLGQRPGQDPTQLGFKINGCVQKSTTTS
REENDREKGENDVVYTNNHHVGSYGTYHNLEHHHHHHQHLSLQADYHSHQLHSLVPPPSQ
ILVCPMTTSPTTTTIQSLFPPSSSSAGSGTMETLDPQM*

>G2140 (148..1254)

ACTCTCTTAACCTTTCTGTTTCTCTCTACCTTCTTTTACCAACCTTTCCTTTCTCTTACA
CACATATATATATACATATATAGAGAGAGAGAAGAGACAAAGAGTTGAAAGATGAAGAC
TCTCATGTCTTCATAGAAACAAGTGATATGTGCGCTAAGAAAGAAGAAGAAGAAGAA
GAAGAAGACAGTTCTGAAGCCATGAACAACATACAAAATTACCAAAATGACCTCTTCTTT
CACCAACTCATCTCTCATCATCACCATCATCATCATGATCCTTCTCAATCTGAAACTTTG
GGAGCATCCGGTAACGTTGGATCTGGTTTCACTATCTTCTCTCAAGATTCCGCTCTCTCCA
ATATGGTCTCTACCTCCACCTACCTCGATCCAACCACCATTTGATCAGTTTCTCTCTCT
TCTTCTTCTCCAGCATCTTTCTACGGAAGTTTCTTCAACAGAAGTCGAGCTCATCATCAG
GGATTACAGTTTGGGTACGAGGGTTTGGTGGAGCCAGTCAGCAGCACATCATCATCAT
GAACAACCTCGGATCTTGTGCGAAGCTTTAGGTCCGGTAGTACAAGCCGGGTCCGGTCCT
TTTGGGTACAAGCTGAGTTAGGGAAGATGACAGCACAAGAGATCATGGACGCTAAAGCT
TTGGCTGCTTCAAAGAGTCATAGTGAAGCTGAGAGAAGAAGAAGAGAGAGAATCAATAAT
CATCTCGCTAAGCTCCGTAGCATATTACCCAACACCACCAAAACGGATAAAGCGTCGTTA
CTAGCTGAAGTGATCCAACATGTGAAAGAGTTGAAGAGAGAGACTTCAGTGATCTCAGAG
ACAAATCTTGTCCCAACGGAAGCGATGAGTTAACGGTAGCTTTCACGGAGGAGGAAGAA
ACCGGAGATGGCAGATTTGTAATTAAAGCGTCGCTTTGCTGTGAAGACAGGTCCGGATCTC
TTGCTTGACATGATTAAACATTGAAAGCTATGCGTCTCAAACGCTCAAGGCGGAGATA
ACCACCGTTGGGGGACGAGTCAAGAACGTTTTGTTTGTACCGGAGAAGAGAGCTCCGGT
GAGGAAGTGGAGGAAGAGTACTGTATAGGGACGATTGAGGAAGCTTTGAAAGCGGTGATG
GAGAAGAGCAATGTAGAGGAATCATCTTCTTCTGAAATGCTAAGAGACAGAGAATGAGT
AGTCACAACACTATCACTATCGTCGAACAACAACAATATAATCAGAGGTAATCAATT
TTTTACTTAAATCGCTTTTTTTTCTTACTTTCGGTGTATCTACTACGTGTGTTGTTGCT
GGTTATGGAATGAATGTTGTACGTACGTTTATCTACTATAGATATATGTGTGTTGTGTGT
ATGTATAACGGAAGTATTTGTATCCGTTGTGCTCTTGGACTTTTGGTTTGGTTCTAAGAT
ACTTATTTTTTAAAACTTGTATCGTTGAGTTGGTTTTCTAGATATGCTTAATGGGAGTAT
GTGACGAAAAAAA

>G2140 Amino Acid Sequence (domain in AA coordinates:167-242)
MCAKKEEEEEEDSSEAMNNIQNYQNDLFFHQLISHHHHHHHDPSQSETLGASGNVGS
FTIFSQDSVSPISLPPPTSIQPPFDQFPFPPSSSPASFYGSFFNRSRAHQGLQFGYEGF
GGATSAHHHHHEQLRLISEALGPVVQAGSGPFLQAEKGMTAQEIMDAKALAASKSHSE
AERRRRERINNHLAKLRSILPNTTKDKASLLAEVIQHVKEKRETSVISETNLVPTESD
ELTVAPTEEEETGDGRFVIKASLCCEDRSDLLPDMIKTLKAMRLKTLKAEITTVGGRVKN
VLFVTGEESSEEEVEEYICIGTIEEALKAVMEKSNVEESSSSGNAKRQRMSSHNTITIVE
QQQQYNQR*

>G2144 (102..1241)

ATTAGGGTTTTGTTGTCGTGAGATTTGATTACACAAATTGCTGAATTTGGTTTCGATTAT
TGGTGTATTGTTTTTCGAAGATTTCCAGTGAGTTTCCGTTTATGGATCTGACTGGAGGAT
TTGGAGCTAGATCCGGCGGTGTTGGACCGTGCCGGGAACCAATAGGCCCTTGAATCGCTAC
ATCTCGGTGACGAATTTCCGGCAACTAGTGACGACTTTACCTCCCGAGAACCCCGCGGTT
CGTTACGGCTTTGCTTGAGCTTCCACCTACACAAGCAGTGAGGCTTCTCCATTTCACTG
ATTCTTCGTCTTCTCAACAAGCGGCAGTGACAGGGATCGGTGGAGAGATTCTCCGCCGC
TTCACTCTTTCCGGTGGGACATTGGCTTTTCTTCTAACTCAGTTCTCATGGAGCGAGCAG
CTCGTTTCTCGGTGATTGCGCACTGAGCAACAAAACGGAATATCTCCGGGGAGACTCCGA
CGAGCTCTGTACCTTCCAATTCAAGTGCTAATCTCGACAGAGTCAAGACGGAGCCTGCTG
AGACCGATTCTCATCTCAGCGGTTGATTTCTGATTACGCGATTGAGAATCAAATCCCTTGCC
CTAACCAGAACAAATCGAAATGGGAAGAGGAAAGATTTCGAAAAGAAGGGTAAAGCTCGA
CGAAGAAGAACAAAAGCTCTGAAGAGAACGAGAAGCTGCCATATGTTTACGTTAGAGCTC
GTCGTGGTCAAGCAACCGATAGCCATAGCTTAGCAGAACGAGCAAGAAGAGAGAAGATAA
ATGCACGAATGAAGCTGTTACAGGAACGGTCCAGGCTGTGATAAGATTCAAGGTACCG
CGCTGGTGTGAGATGAAATCAATACCATGTCCAGTCATTACAACGTCAAGTGGAGATGC
TATCAATGAGACTTGCTGCGGTAAACCCAGAAATCGACTTCAATCTCGACACCATATTGG

CTTCAGAAAACGGTTCTTTAATGGATGGGAGCTTCAATGCCGCACCAATGCAGCTTGCTT
GGCCTCAGCAAGCCATTGAGACCGAACAGTCCTTTTCATCACCAGGCAACTGCAACAACCAC
CAACACAACAATGGCCCTTTTGACGGCTTGAACCAGCCGGTATGGGGAAGAGAAGAGGATC
AAGCTCATGGCAATGATAACAGCAATTTGATGGCAGTTTCTGAAAATGTAATGGTGGCTT
CTGCTAATTTGCACCCAAATCAGGTCAAATGGAGCTGTAAGTTGGGAAAACGGTAGAGA
TCATGAATGTGTATATACATCGTATAAGCTCGTTTCTCTATATAAATATAATCATAAA
TATAGATATCTGTTAAGAAGGTATCAGTCATTTGATTTCAGAGAGACAACACTGGTATGAT
TGTTTCTTATTTCTGTACCAGATTTTCGACAATGTAGAATTTAGTAGGATATGATCATTTT
GATCTCGTTATATATA

>G2144 Amino Acid Sequence (domain in AA coordinates:203-283)
MDLTGGFGARSGVGPCREPIGLES LHLGDEFRLVTTLPENPGGSFTALLELPPTQAV
ELLHFTDSSSSQAAVTGIGGEIPPLHSGGTLAFPSNSVLMERAARFSVIAEQONGN
ISGETPTSSVPSNSSANLDRVKTEPAETDSSQRLISDSAIENQIPCPNQNNRNGKRKDFE
KKGKSSTKKNKSSEENEKLPYVHVRARRGQATDSSHSLAERARREKINARMKLLQELVPGC
DKIQGTALVLDEIINHVSQSLQRQVEMLSMRLAAVNPRIDFNLDITILASENGSLMDGSFNA
APMQLAWPQQAIETEQSFHHRQLQOPPTQQWPFGLNQPVWGREEDQAHGNDNSNLMAVS
ENVMVASANLHPNQVKMEL*

>G2431 (47..1057)
CCCTTTCGTTTTTATTTAAATTTCTTGGGTCGTTTCTTAAATTTGTATGTGTTTATTAAT
GGAGATCAACAATAATGCCAACATACTAATACTACTATTGATAATCACAAGGCAAAGAT
GAGCCTTGTTGTGTCAACGGATGCTAAGCCAAGGTGAAATGGACTTGTGATCTTCATCA
CAAATTCATCGAAGCCGTTAATCAACTTGGAGGACCTAACAAAGCAACACCTAAGGGTTT
GATGAAGGTTATGGAGATTCTCGGCTTACCTTATACCATCTCAAGAGCCATTACAGAA
ATATCGGTTAGGGAAGAGCATGAAGTTCGATGATAACAAGCTAGAAGTTTCTCTGCATC
AGAGAATCAAGAAGTTGAGAGTAAAAACGATTCAAGAGATCTCCGAGGCTGCAGTGTCAC
CGAAGAAAACAGCAATCCAGCTAAAGAAGGGCTACAAATCACAGAGGCTTTACAAATGCA
GATGGAAGTTTCAGAAGAACTTCATGAACAAATCGAAGTTCAGAGGCATTTGCAGGTGAA
GATTGAGGCACAAGGAAAGTATCTACAGTCCGTTTAAATGAAAGCTCAACAACTCTCGC
TGGCTACTCATCTTCAAATCTCGGCATGGATTTTGGCAGGACCGAGCTCTCTAGATTAGC
TTCATGGTGGAACAGAGGCTGTCCAAGCACTTCGTTCTCAGAGCTAACGCAAGTAGAAGA
AGAAGAAGAAGGTTTCTTGTGTGTACAAGAAACAGAAACAGAGGAATTAGTCAGCTGAG
ATGTTCAGTAGAGAGCTCGTTGACATCTTCAGAGACCTCAGAGACAAAACGGATACTGA
CAATAACCTTAATAAATCGATTGAACCTCCGTTGATGGAGATCAACTCGGAAGTGATGAA
GGGGAAGAAGAGAAGCATAAACGACGTCGTTTGGCTGGAGCAGCCTCTAATGAAGAGAGC
TTTTGGAGTTGATGATGATGAGCATTTGAAGTTGAGTTTGAATACTTACAAGAAAGACAT
GGAGGCGGTGTACGAACATAGGACTAGGGTTTAATTAATAAAAAAAACATTTTACTAAAGTT
ATATAAAAATGTTTAAAGAAATCCA

>G2431 Amino Acid Sequence (conserved domain in AA coordinates:38-88)
MCLLMEINNANNTNTTIDNHKAKMSLVLSTDAKPRCLKWTCDLHHKFIEAVNQLGPNKA
TPKGLMKVMEIPGLTLYHLKSHLQKYRLGKSMKFDDNKLEVSSASENQEVESKNDSRDLR
GCSVTEENSNPKEGLQITEALQMOMEVQKKLHEQIEVQRHLQVKIEAQGKYLQSVLMKA
QOTLAGYSSSNLGMDFARTELSRLASMVNRGCPSTSFSELTQVEEEEGFLWYKKPENRG
ISQLRCSVESSLTSSETSETKLDTDNLNKSIELPLMEINSEVMKGKKRSINDVVCVEQP
LMKRAFGVDDDEHLKLSLNTYKDM EACTNIGLGFN*

>G2465 (86..1150)
CAATATTCTTTCTCCATTGAGATTAAGCTTCTTTCTCGCTGTCTCTCTATAGATCTT
GGTCTTAGTCCCTTTTGAATAATAATGATGGTGGAGATGGATTACGCTAAGAAAATGCA
GAAATGTGATGAATACGTTGAAGCACTTGAAGAAGAACAGAGAAAATCCAAGTCTTTCA
ACGCGAGCTTCCCTTTATGTTTAGAGCTTGTCACTCAAGCGATCGAAGCTTGTCCGAAGGA
GTTATCTGGTACGACGCAACTACATCAGAACAGTGTTCAGAACAGACCACAAGTGTG
TGGTGGTCTGTCTTTGAAGAGTTTATTCTTATCAAGAAAATTAGTTCCTTGTGTGAAGA
AGTACAAGAAGAAGAAGAAGATGGTGAACATGAATCTTCTCCAGAACTTGTGAATAA
TAAGAAATCAGATTGGCTTAGATCTGTTTCAAGCTATGGAATCATTACCGGATCTAAATCC
AAAAGAGGAGCGTGTAGCTAAGAAAGCGAAAGTGGTGGAGGTGAAACAAAAAGCGGTGC
GTTTCAGCCGTTTCAAAGCGCGTTTGGAGACTGATTTGCAACCGGCGGTGAAAGTAGC
TAGTTGATGCCAGCGACGACGACGAGTTCACGACGGAACCTTGTGGTGGTAAAGTGA
TTTGATTAAAGCTGGAGATGAGGAAAGACGGATAGAGCAGCAGCAATCGCAGTCGCATAC

GCATAGAAAACAAAGGCGGTGCTGGTCGCCGAATTACACCGTCGATTCTTAAACGCGCT
 TCAGCAGCTTGGAGGATCTCATGTTGCTACACCAAAGCAAATCAGGGATCACATGAAGGT
 TGATGGATTAAACAAACGACGAAGTTAAAAGCCATTTACAGAAATATAGACTTCACACAAG
 AAGGCCAGCAGCAACATCCGTGGCGGCACAAAGTACCGGGAATCAGCAACACCACAATT
 TGTGGTGGTTGGAGGCATATGGGTACCATCGTCACAAGATTTTCCACCACCGTCCGATGT
 AGCCAACAAGGTTGGTGTATATGCTCCGGTTGCGGTGGCGCAATCTCCAAAACGTTTCGTT
 GGAGAGAAGTTGCAACTCGCCGGCGGCATCTTCTCTACAAATACAAATACTTCTACTCC
 TGTGTATAATCTGATAGTCATACTATAATCATCTCTGATGTTGATTTTGGTGTAGGTT
 TGAAAATGTTTATGTGAATGTAA

>G2465 Amino Acid Sequence (conserved domain in AA coordinates:219-269)

MMVEMDYAKMKQKCHEYVEALEEEQKKIQVFQRELPLCLELVTQAIACRKELSGTTTTT
 SEQCSEQTTSVCGGPVFEEFIPIKKISSLCEEVQEEEEEDGEHESSELPVNNKSDWLRS
 VQLWNHSPDLNPKERVAKKAVVEVKPKSGAFQFPQKRVLETDLPVAVKVASMPATTT
 SSTTETCGGKSDLIKAGDEERRIEQQSQSHTRKQRRCSPELHRRFLNALQQLGGS HV
 ATPKQIRDHMKVDLNDDEVKSHLQKYLHTRRPAATSVAAQSTGNQQPQFVVVGGI WV
 PSSQDFPPPSDVANKGGVYAPVAVAQSPKRSLESCNSPAASSSTNTNTSTPVS*

>G2583 (38..607)

CAAATCAGAAAATATAGAGTTTGAAGGAAACTAAAAGATGGTACATTGAGGAAGTTCCG
 AGGTGTCCGCCAGCGACAATGGGGTCTTGGGTCTCTGAGATTGCCCATCTCTATTGAA
 GAGAAGAGTGTGGCTTGGAACCTTCGAAACGGCAGAAGCGGCTGCAAGAGCATACGACCA
 AGCGGCTCTTCTAATGAACGGCCAAAACGCTAAGACCAATTTCCCTGTCGTAATAATCAGA
 GGAAGGCTCCGATCACGTTAAAGATGTTAACTCTCCGTTGATGTCACCAAGTCATTATC
 TGAGCTTTTGAACGCTAAGCTAAGGAAGAGCTGCAAGACCTAACGCCTTCTTTGACGTG
 TCTCCGTCTTGATACTGACAGTTCCACATTTGGAGTTTGGCAGAAACGGGCCGGGTGCAA
 AACAAGTCCGACTTGGGTTCATGCGCTCGAAGTTGGGAACGTAGTCAACGAAAGTGGCGT
 TGACTTAGGGTTGACTACGATGAACAAACAAAACGTTGAGAAAGAAGAAGAAGAAGA
 AGCTATTATTAGTGATGAGGATCAGTTAGCTATGGAGATGATCGAGGAGTTGCTGAATTG
 GAGTTGACTTTTGACTTTAACTTGTGCAAGTCCACAAGGGGTAAGGGTTTTTC

>G2583 Amino Acid Sequence (domain in AA coordinates:4-71)

MVHSRKFRGVRQRQWGSWVSEIRHPLLKRRVWLGTFFETAEEAARAYDQAALLMNGQNAKT
 NFPVVKSEEGSDHVKDVNSPLMSPKLSSELLNAKLKSKDLTPSLTCLRLDTS SHIGV
 WQKRAGSKTSPTWVMRLELGNVNVESAVDLGLTMMNKQNVKEEEEEEEAIISDEDQLAME
 MIEELLNWS*

>G2724 (1..651)

ATGGAATAGAAATAAGGAGAGGTCCATGGACTGTGGAAGAAGACATGAAGCTCGTCAGT
 TACATTTCTCTTACCGGTGAAGGAGATGGAACCCCTCTCTCGTTCTGCTGGACTGAAT
 AGAACGGGGAAAAGTTGCAGATTGCGGTGGCTAAATTATCTCCGGCCGGATATCCGCCGT
 GGAGACATATCCCTTCAAGAACAATTTATCATCCTTGAATCCATTCTCGTTGGGGAAAT
 CGGTGGTCAAAGATTGCTCAACATTTACCGGGAAGAAGACAGATAACGAGATAAAGAATTAT
 TGAGAAACACGTGTTCAAAAGCATGCAAAACTTCTAAATGTGACGTGAACAGCAAGCAA
 TTCAAAGACACCATCAAAATCTCTGGATGCCTCGTCTCATCGAGAGAATCGCCGCCACT
 CAAAGTGTCCAATTTACCTCTAACCCTACTCGCTGAGAACTCCAGCGTCGCCACCGCC
 ACGTCATCAACGTCGTCGTGAGGCTGTGAGATCGAGTTTCTACGGTGGTGATCAGGTG
 GAATTTGGAACGTTGGATCATATGACAAATGGTGGTTATTGGTTCAACGGCGGAGATACG
 TTTGAAACTTTGTGTAGTTTTGACGAGCTCAACAAGTGGCTCATACAGTAG

>G2724 Amino Acid Sequence (conserved domain in AA coordinates:7-113)

MEIEIRRGPTVEEDMKLVSYISLHGEGRWNSLSRSAGLNRTGKSCRLRWLNLYLRPDIRR
 GDISLQEQFIILELSRWGNRWSKIAQHLPGRDNEIKNYWRTRVQKHAKLLKCDVNSKQ
 FKDTIKHLWMPRLIBERIAATQSVQFTSNHYPENSSVATATSSSTSSSEAVRSSFYGGDQV
 EFGTLDHMTNGGYWNGGDTFETLCSFDELNKWLIQ*

>G377 (1..396)

atgggtctctcgcattttccaacagcgctcagaaggagtactaccacttctggtgatgaac
 acggttggtttcaatcactctgttgaagaacatgggtgaggtctgttttcaaattgttgca
 tccgagactgaatctccatggagatagacgacgagcctgaagatgattttgtactaga
 agaactctcgataacacagttcaagtctctatgtgagaacatagaagaggaagaagaagag
 aaaggtgtggagtgtgtgtgtgcctttgtgggttaaagaggaagaggaagtgagtga
 ttggtttcttgcaagcatttctccacagagcctgtctagacaactgggttgtaataac

cacaccacatgccctctttgcaggtccattctctag

>G377 Amino Acid Sequence (domain in AA coordinates: 85-128)

MGLSHFPTASEGVLPPLVMNTVVSITLLKNMVRSVFQIVASETESSMEIDDEPEDDFVTR
RISITQFKSLCENIEEEEEKGVCECCVCLCGFKEEEVSELVSKHFFHRACLDNWFGNN
HTTCPLCRSIL*

>G428 (97..1032)

TTACTTTTGTGTTTCTTCATATTCTTCAGAAGCAAGCACAAGGCTAGGGATCGAAGAAGC
GGCGATCACTGATCGTATCTCACTACGATCACATTAATGGATAGAATGTGTGGTTTCCGC
TCGACGGAAGACTATTCGGAGAAAGCGACGTTGATGATGCCGTCCGATTATCAGTCTTTG
ATTTGTTCAACCACCGAGACAATCAAAGACTGTTTGGATCCGACGAACTCGCTACCGCT
TTGTCTCTCGGAGTTGCTTCCGCGTATTTCGAAAAGCTGAGGATAATTTCTCTCTTAGTGTC
ATCAAATCCAAATCGCTTCTCATCTTTGTATCCTCGCTTACTCCAAACCTACATCGAT
TGCCAAAAGGTGGGAGCGCTTATGGAAATAGCGTGTATATTGGAAGAGATTGAGCGAGAG
AACCATGTGTACAAGAGAGATGTTGCTCCATTATCTTGCTTTGGAGCTGATCCTGAGCTT
GATGAATTCATGGAAACCTACTGTGATATATTGGTTAAATACAAACCGATCTTGCGAGG
CCGTTTCGACGAGGCTACAACTTTCATAAACAAGATTGAAATGCAGCTTCAGAACTTGTGC
ACTGGTCCAGCGTCTGCTACAGCTCTTTCAGATGATGGTGCGGTTTCATCTGACGAGGAA
CTGAGAGAAGATGATGACATAGCAGCGGATGACAGCCAACAAAGAAGCAATGACCGCGAT
CTGAAGGACCAGCTACTACGCAAATTTGGTAGCCATATCAGTTTCAATGAAACTCGAGTTC
TCTAAAAAGAAGAAGAAAGGAAGCTACCAAGAGAAGCAAGACAAGCGTTGCTCGATTGG
TGGAATGTTTCATAATAAATGGCCTTACCCTACTGAAGGCGACAAAATAGCTCTGGCTGAA
GAAACAGGTTTGGATCAAAAACAAATCAACAATTGGTTTATAAACCAAGGAAACGCCAT
TGGAAGCCTTCGGAGAACATGCCGTTTGATATGATGGACGATTCTAATGAAACATTCTTT
ACCGAGGAATGAAAAGAGAGACATGGGATTGTGCAATTGTATAATTTTACACTGTTTTCC
CAAGAAAAGAAAACAGTAAAAGCTTTTGGTAAATGGGACATCATCGCAATGAATGGAA
CCAGTTAGCCAAAACGGTCAAGGGCGTGGCGTAACGAGACATTGTATTGGAAATAGTGGC
AATATTATGTCACTAATCTTCCAATGGTCCAAAATGATAGATTCTTATTTGTATTGAAC
CTTACTTAGATAGCTGATGTGTCACTAAATAATTTATTTTCATCCTTATACTACTTGTA
TCAATGTCTCTAATTGATCAATTGTTGCTTGCTATTCAAAAAAAAAAAAAAAAAAAAAA

>G428 Amino Acid Sequence (domain in AA coordinates: 229-292)

MDRMCGFRSTEDYSEKATLMMPDYSQSLICSTTGDNRQLFGSDELATALSSELLPRIRKA
EDNFSLSVIKSKIASHPYLRLLQTYIDCQKVGAPMEIACILEIQRENHVKRDRVAPLS
CFGADPELDEFMETYCDILVKYKTDLARPFDEATTFINKIEMQLQNLCTGPASATALSDD
GAVSSDEELREDDIAADDSQQRSDNRDLKDQLLRKFGSHISSLKLEFSKKKKKGKLPRE
ARQALLDWNVHNKWPYPTEGDKIALAETGLDQKQINNWFINQRKRHWKPSNMFPDMM
DDSNETFFTEE*

>G447 (241..3501)

CTTTTAAAGAGCTTAAAAATTTGCTTTGAAGCTTCAAATATTCTTATGAACTAAAAAGAA
GAAAAAGCTTTTGTCTTTTCTTTTCTTAGCAGCAGAATGATTTTGTCTTCAAAATATT
ACTATTTAGTTTCTCTCGTGCTCTTCTCTTGAGCAAATACAGATTCGTTAATTTTGCTGA
AGAAGAAGAACTCTGTTCTTCCCTGCACCAAACCAATTTTTCGTTCTTTCTATAAACC
ATGAAAGCTCCATCAAATGGATTTCTTCCAAGTTCCAACGAAGGAGAGAAGAAGCCAATC
AATTTCAACTATGGCAGCGTTGTGCGAGGGCTTTAGTTTCATTACCTCCTGTGGGAAGT
CTTGTGGTTTACTTCCCTCAAGGACACAGCGAGCAAGTTGCAGCATCGATGCAGAAGCAA
ACAGATTTTATACCAAATTACCCAAATCTTCTTCTAAGCTGATTTGCTTGCTTCACAGT
GTTACATTACATGCTGATACCGAAACAGATGAAGTCTATGCACAAATGACTCTTCAACCT
GTGAATAAGTATGATAGAGAAGCATTGCTAGCTTCTGATATGGGCTTGAAGCTAAACAGA
CAACCTACTGAGTTTCTTGAAGACTCTTACTGCAAGTGACACAAGCACTCATGGTGGA
TTCTCTGTACCGCGTCGTGCAGCTGAGAAAATATCCCTCCTCTTGATTTCTCGATGCAA
CCGCTGCGCAAGAGATTGTAGCTAAAGATTTACATGATACTACATGGACTTTTCAACAT
ATCTATCGAGGCCAACCAAAAGACACTTGCTTACCACAGGTTGGAGCGTTTGTGTTAGC
ACAAAGAGACTATTTGCGGGTGATTCAGTTTGTGTTGTAAGAGATGAGAAATCACAGCTG
ATGTTGGGTATAAGACGTGCAATAGACAAACTCCGACTCTTCCCTCATCGGTCATATCC
AGCGACAGTATGCACATTGGGATACTTGCAGCTGCAGCTCATGCTAATGCCAATAGTAGC
CCTTTTACCATCTTCTTCAATCCAAGGGCAAGTCTTCCAGAGTTGTAGTTCTTTAGCC
AAATACAACAAAGCCTTATACGCTCAAGTATCTTAGGAATGAGATTCGGGATGATGTTT
GAGACTGAGGATTGTGGGGTTTCGTAGATATATGGGTACAGTCACAGGTATTAGTGATCTT

GACCTGTAAAGATGGAAAGGCTCACAATGGCGTAATCTTCAGGTAGGATGGGATGAATCA
ACAGCTGGAGATAGGCCAAGCCGAGTATCCATATGGGAAATCGAACCCTCATAACTCCT
TTTACATATGTCTCCTCCATTTTTCAGACCTAAGTACCCGAGGCAACCCGGGATGCCA
GATGATGAGTTAGACATGGAAAAATGCTTTCAAAAGAGCAATGCCTTGGATGGGAGAAGAC
TTTGGGATGAAGGACGCACAGAGTTCGATGTTCCCTGGTTTAAAGTCTAGTTCAATGGATG
AGTATGCAGCAAAACAATCCATTGTTCAGGTTCTGCTACTCCTCAGCTCCCGTCCGCGCTC
TCATCTTTTAACTTACCAACAATTTTGCTTCCAACGACCCCTCCAAGCTGTTGAACCTC
CAATCCCCAAACCTCTCTCCGCAAAATCCCAATTCAACAAACCGAACACGGTTAACCAT
ATCAGCCAACAGATGCAAGCACAACCAGCCATGGTGAAATCTCAACAACAACAACAACA
CAACAACAACAACCAACACCAACAACAACAACCTGCAACAACAACAACAACCTACAGATG
TCACAGCAACAGGTGCAGCAACAAGGGATTTATAACAATGGTACGATTGCTGTTGCTAAC
CAAGTCTCTTGTCAAAGTCCAAACCAACCTACTGGATTCTCTCAGTCTCAGCTTCAGCAG
CAGTCAATGCCTCCTACTGGTGTCTAAAATGACACACCAGAACATAAATCTATGGGGAAT
AAAGGCTTGTCTCAAATGACATCGTTTGGCGAAGAAATGCAGTTTCAGCAGCAACTGGAA
ATGCATAACAGTAGCCAGTTATTAAGAAACCGCAAGAAGTCTCTCTCCATTCTTA
CAACAAAATCTGTCCCAAAATCCTCAGCAACTCCAAATGCAACAACAATCATCAAAACCA
AGTCTTTCACAACAGCTTCAGTTGCAGCTACTGCAGAAGCTACAGCAGCAGCAACAGCAG
CAGTCGATTCTCCAGTAAGTCTATCCTTACAGCCACAATTATCAGCGTTGCAGCAGACA
CAAAGCCATCAATTGCAACAACCTTCTGTCTCTCAAAATCAACAGCCCTTGGCAGATGGT
AATAACAGCTTCCCAGCTTCACTTTTCATGCAGCCTCCACAGATTGAGGTGAGTCTCAG
CAGCAAGGACAGATGAGTAACAAAAATCTGTAGCCGCTGGAAGATCACATTCTGGCCAC
ACAGATGGAGAAGCTCCTTCTTGTTCAAACCTCACCTTCCGCCAATAACACGGGACATGAT
AATGTTTTCACCGACAAATTTCTGAGCAGAAATCAACAGCAAGGACAAGCTGCATCTGTA
TCTGCATCTGATTAGTCTTTGAGCGCGCAAGCAATCCGGTCCAAGAGCTTTATACAAAA
ACTGAGAGCCGGATCAGTCAAGGCATGATGAATATGAAGAGTGTGGTGAACATTTCAGA
TTTAAAGCGCGGTAAACAGATCAAAATCGATGTATCCACAGCGGGAACGACGTACTGTCT
GATGTTGTTGGCCCTGTACAGCAGCAACAACTTTCCCACTACCATCATTGTTGTTGAT
GGAGATGCCAATCTCATCATCAAGAAACAACCTTAGCTTTCCCTGGTAATCTCGAAGCC
GTAACCTTCTGATCCCTCTTCTCAAAAGGACTTTCAAACTTGGTTCCCAACTATGGC
AACACACCAAGAGACATTGAGACGGAGCTGTCCAGTGCTGCAATCAGTTCTCAGTCATTT
GGTATTCCCAGCATTCCCTTTAAGCCCGGATGTTCAAATGAGGTTGGCGGCATCAATGAT
TCAGGAATCATGAATGGTGGAGGACTGTGGCCCAATCAGACTCAACGAATGCGAACATAT
ACAAAGGTTCAAAACGAGGGTCAGTAGGTAGATCAATAGATGTTACCCGTTATAGCGGC
TATGATGAACCTTAGGCATGACTTAGCGAGAATGTTTGGCATCGAAGGACAGCTCGAAGAT
CCGCTAACCTCTGATTGGAACTCGTCTACACCGATCAGCAAAACGATATTTTACTAGTT
GGTGTATGATCCTTGGGAAGAGTTTGTGAAGTGCCTGCAGAACATAAAGATACTATCATCA
GTAGAAGTTCAGCAATGAGCTTAGACGGAGATCTTGCAGCTATCCCAACCACAAACCA
GCCTGCAGCGAAACAGACAGCGGAAATGCTTGGAAAGTACACTATGAAGACACTTCTGCT
GCAGCTTCTTTCAACAGATAGAAATAAAAAGATGCAATATACCAAGTCAACTTACATTA
TCATTGAGGCCATCGCAAAGTACATGTTTCTTTTGTGTGTATGTACTGCAACAACAA
ACTGAGAAGAAGAAGATACTGCACGGTATATAAACATTTTATAGGACAGTGATTGATT
TTTCATTCTAAGTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT
TGCTTGACAAGTCTATGAGGAGCATATCTTATACAGAGATACTAAGATGTAATGTTAATG
TAACTAAACAATTACCTTCATTAATCATGAATCCTTTGGTCTGTTTAAAA

>G447 Amino Acid Sequence (conserved domain in AA coordinates:22-356)

MKAPSNGLPSSNEGEKKPINSQWLWHACAGPLVSLPPVGSVYVFPQGHSEQVAASMQKQ
TDFIPNYPNLPKSLICLLHSVTLHADTETDEVYAQMTLQPVNKYDREALLASDMLKLN
QPTTEFFCKTLTASDSTHGGFSVPRRAEKKIFPPLDFSMQPPAQEIVAKDLHDTTWTFRH
IYRGQPKRHLTLTGWSVVFSTKRLFAGDSVLFVRDEKSQLMLGIRANRQTPTLSSSVIS
SDSMHIGILAAAAHANANVSPTIFFNPRASPSEFVVPLAKYNKALYAQVSLGMRFRMMF
ETEDCGVRRYMGTVTGISDLDPVRWKGSQWRNLQVGWDESTAGDRPSRVSWEIEPVITP
FYICPPPPFRPKYRQPGMPDDELDMENAFKRAMPWMGEDFGMKDAQSSMFPGLSLVQWM
SMQNNPLSGSATPQLPSALSSFNLPNNFASNDPSKLLNFQSPNLSSANSQFNKPNVTNH
ISQQMQAPAMVKSQQQQQQQQQHQHQQQQLQQQQQLQMSQQQVQQQGIYNNGTIAVAN
QVSCQSPNQPTGFSQSLQQQSMPLTGAKMTHQNINSMGNKGLSQMTSFAQEMQFQQQLE
MHNSSQLLRNQEQSSLHSLQQLNSQNPQQLQMQQSSKPSPSQQLQLQLLQKLQQQQQQ
QSIPPVSSSLQPLSALQQTQSHQLQLLSSQNQQPLAHGNNSFPASTFMQPPQIQVSPQ

QQGQMSNKNLVAAGRSHSGHTDGEAPSCSTSPSANNTGHDNVSPTNFLSRNQQQGQAASV
 SASDSVFERASNPVQELYTKTESRISQGMNMKSAGEHFRFKSAVTDQIDVSTAGTTYCP
 DVVGPVQQQTFPLPSFGFDGDCQSHHPRNNLAFPGNLEAVTSDPLYSQKDFQNLVPNYG
 NTPRDIETELSSAAISSQSGFIPSIIPKPGCSNEVGGINDSGIMNGGLWPNQTQRMRTY
 TKVQKRGSVGRSIDVTRYSGYDELRLHDLARMFGIEGQLEDPLTSDWKLVTYDTHENDILLV
 GDDPWEEFVNCVQNIKILSSVEVQMSLDGDLAAIPTTNQACSETDSGNAWKVHYEDTSA
 AASFNR*

>G464 (41..760)

CTCTGCTGGTATCATTGGAGTCTAGGGTTTGTATTGACATGCGTGGTGTGTCAGAATT
 GGAGGTGGGGAAGAGTAATCTTCCGGCGGAGAGTGAGCTGGAATTGGGATTAGGGCTCAG
 CCTCGGTGGTGGCGCGTGGAAAGAGCGTGGGAGGATTCTTACTGCTAAGGATTTTCCTTC
 CGTTGGGTCTAAACGCTCTGCTGAATCTTCTCTCACCAGGAGCTTCTCTCTCGTTC
 AAGTCAAGTGGTAGGATGGCCACCAATTGGGTTACACAGGATGAACAGTTTGGTTAATAA
 CCAAGCTATGAAGGCAGCAAGAGCGGAAGAAGGAGACGGGAGAAGAAAGTTGTGAAGAA
 TGATGAGCTCAAAGATGTGTCAATGAAGGTGAATCCGAAAGTTCAGGGCTTAGGGTTTGT
 TAAGTGAATATGGATGGAGTTGGTATAGGCAGAAAAGTGGATATGAGAGCTCATTCTGTC
 TTACGAAAACCTTGCTCAGACGCTTGAGGAAATGTTCTTTGGAATGACAGGTACTACTTG
 TCGAGAAAAGGTTAAACCTTTAAGGCTTTTAGATGGATCATCAGACTTTGTACTCACTTA
 TGAAGATAAGGAAGGGGATTGGATGCTTGTGGAGATGTTCCATGGAGAATGTTTATCAA
 CTCGGTGAAAAGGCTTCGGATCATGGGAACCTCAGAAGCTAGTGGACTAGCTCCAAGACG
 TCAAGAGCAGAAGGATAGACAAAGAAAACACCTGTTTAGCTTCCCTTCAAAGCTGGCA
 TTGTTTATGTATTGTTGAGGTTTGAATTTACTCGATACTTTTGAAGAAAGTATTTTG
 GAGAATATGGATAAAAGCATGCAGAAGCTTAGATATGATTTGAATCCGGTTTTTCGGATAT
 GGTTTTGCTTAGGTCATTCAATTCGTAGTTTTCCAGTTTGTCTTCTTTGGCTGTGTAC
 CAATTATCTATGTTCTGTGAGAGAAAGCTCTT

>G464 Amino Acid Sequence (domain in AA coordinates: 20-28, 71-82, 126-142, 187-224)

MRGVSELEVGSNLPASELELGLGLSLGGGAWKERGRILTAKDFPSVGSKRSAESSSHQ
 GASPPRSSQVVGWPPIGLHRMNSLVNNQAMKAARABEGDGEKKVVKNDLKDVSMTKVNPK
 VQGLGFVKVNMDDGVGIGRKVDMRAHSSYENLAQTLLEEMFFGMTGTTTCREKVKPLRLLDGS
 SDFVLTYYEDKEGDWMLVGDPWRMFINSVKRLRIMGTSEASGLAPRRQEQKDRQRNPNV*

>G557 (192..698)

CAGAGATCTGACGGCGGTAGCAGAGTAATCTATTCTCTCCCAAATGTCTCGCAATTAGA
 TTCTTTTCCAAGTTCTTCTGTAAATCCCAAGTCCCGCTCTTTTCTCTTTATCTTTTAC
 CAGCTTCGCTACTAAGACAACAAATCTTCCCTCTCTCTCTCGCCTGATCGATCTTCAA
 GAGTAAGAAAAATGCAGGAACAAGCGACTAGCTCTTTAGCTGCAAGCTCTTTACCATCAA
 GCAGCGAGAGGTATCAAGCTCTGCTCCACATTTGGAGATCAAAGAAGGAATTGAAAGCG
 ATGAGGAGATACGGCGAGTGCCGGAGTTTGGAGGAGAAGCTGTGCGAAAAGAACTTCCG
 GTAGAGAATCTGGATCGGCGACCGGTCAGGAGCGGACACAGGCGACTGTGCGAGAAAGTC
 AAAGGAAGCGAGGGAGGACACCGCGGAGAAAGAGAACAAGCGGCTGAAGAGGTTGTTGA
 GGAACAGAGTTTTCAGCTCAGCTCAGCAAGCAAGAGAGAGGAAAAAGGCTTACTTGAGCGAGTTGG
 AAAACAGAGTGAAAGACTTGGAGAACAAAACTCTGAACCTGAAGAGCGACTCTCTACTC
 TTCAGAACGAGAACCAGATGCTTAGACATATTCTGAAGAACACAACAGGAAACAAGAGAG
 GAGGTGGTGGTGGTTCTAATGCTGATGCAAGCCTTTGATCTCCTTCTTCTTGTGTTA
 TATTTTTGTGGATAAAATTTACAGAGAATTGTATCAATAATTATCATGTTAAAAATTATAT
 GGGATGTGAGAGCTAATATTGCAATTGTAGACCAAGTTCTCTTAAAAAATAAAAAA
 AA

>G557 Amino Acid Sequence (domain in AA coordinates: 90-150)
 MQEQATSSLAASSLPSSSSSSAPHLIEKIEGIESDEIRRVPEFGGEAVGKETSGRES
 GSATQOERTQATVGESQQRKGRTPAEKENKRLKRLLRNRVSAQQARERKKAYLSELENRV
 KDLENKNSELEERLSTLQENQMLRHILKNTTGNKRGGGGSNADASL*

>G577 (44..2155)

AAAAACAGACTGAGAGAGAGAGAGAGAGTGTGTTGTTGGCCATGGGATGCACGGCCTC
 CAAGCTCGACAGTGAAGATGCTGCTCCGTCGCTGCAAGGAGCGCGCGCTTATGAAGGA
 CGCGTCTACGCTCGTCAACATCTCGCCGCCGCTCACTCTGACTACTGCCGCTCCCTTCG
 TCTCACTGGCTCTGCCCTCTCTCTCTCGCCGCCGCGAGCCCTCTCCGCTCTCCGAGAA
 TACTCCCGCTGTTTTCTCCGCCCTTCTCCAGTCAGGACGCGCCACGTGTCCCTTCTTC

CCATTCCCCAGAACCCCTCCTCCGCCATCCGCAGCAAGCCTAAGCCTACTAGGCCTAG
 GAGGCTTCCACACATTCTCTCCGACTCCTCTCCTTCTTCTCTCCTGCCACCAGTTTCTA
 TCCCCTGCTCACCAGAACTCTACTTACTCTCGCTCTCCATCTCAAGCTTCTCTGTCTG
 GAACTGGGAGAATTTCTACCTCCTCTCCCCCGACTCCGAGTACTTCGAACGCAAAGC
 TCGCCAGAACCACAAGCACCGTCTCTCCGACTACGACGCCGAAACTGAAAGATCCGA
 CCACGATTACTGCCACTCACGGAGAGATGCCGCCGAGGAAGTTCACTGCAGCGAGTGGGG
 CGACGACCACGACCGTTTCACTGCCACCTCTCTCGTCCGACGGAGATGGGGAGGTGAAAC
 TCACGTTTCCAGATCCGGTATTGAAGAAGAGCCTGTGAACAACACATCAAGACCCAAA
 TGGCAAAGAGCACTCTGACCATGTTACCCTTCTCCGACTGCTACAAGACCAAATTGGT
 GGTAAAGCACAAGAATTTGAAGGAGATCCTTGACGCCGTTCAAGACTACTTCGACAAGGC
 TGCCTCCGCTGGGGACAGGTCTCCGCCATGCTTGAGATCGGCCGGGTGAGCTCGACCG
 CAGCTTCAGCAAGCTGAGGAAGACGGTGTATCATTCAAGCAGTGTGTTAGCAACTTGAG
 CGCAAGCTGGACCTCAAAACCCCATTTGGCAGTCAAATACAAGCTCGATGCATCTACCTT
 GAATGATGAACAAGGCCGGCTCAAGAGCCTTCTGCTCCACTCTAGACCGACTCCTCGCTTG
 GGAAGAAGCTTTATGAGGATGTCAAGGCAAGAGAAGGAGTTAAGATTGAGCACGAGAA
 GAAGCTGTCTGCGCTGCAGAGTCAGGAGTATAAGGGAGGTGATGAATCCAAGCTAGACAA
 GACTAAACTTCCATAACAGACTGCAATCACTCATCATTGTTTCTTCAGAAGCTGTTTT
 AACCACGTCTAATGCCATTCTCCGCCCTCCGGGACACTGACCTTGTCCTCAGCTTGTTGA
 ACTCTGCCACGGATTAATGTACATGTGGAAGTCAATGCACGAGTATCAGGAAATCCAGAA
 CAACATCGTGCAACAAGTCCGTGGCTGTATCAACCAACAGAGAGAGGTGAGTCAACATC
 AGAGGTACACCGGACGGTGACGCGGACCTAGAGTCAGCTGTGTCCTTGTTGGCATTTCGAG
 CTTCTGTGCGCATCATTAATTCAGAGGGAGTTCATATGCTCTCTCCACGATGGTTCAA
 GCTGAGCCTGGTTCCCTGAGCAACGGAGACCCAAAGAAACAGCGGCCAGACTCATTTCG
 CTTGTGCGAGGAGTGGAAGCAGAGCCTGGAACGGGTGCCTGACACAGTGGCGTCAGAAGC
 CATAAGAGCTTTGTAAACGTGGTACATGTGATATCAATAAGCAGGCGGAAGAGGTGAA
 GATGAAGAAACGCACGGAGAGTGCAGGAAAGGAGCTGGAGAAGAAAGCATCCTCAGTGA
 GAGCATAGAGAGGAGTACTACAGGCATACTCGACGGTTGGGATAGGCCCTGGACCGGA
 GGTGTTGGACTCACGGGACCCGCTATCTGAGAAGAAATGTGAGCTGGCGGCATGTCAGAG
 GCAGGTGGAGGATGAGGTAATGAGGCACGTGAAGGCTGTGGAGGTGACACGAGCTATGAC
 TCTCAACAATCTACAAACCGGCTGCCCAATGTATTCCAGGCCTTGACCAGCTTCTCATC
 TCTCTTCACTGAATCTCTCCAGACTGTCTGTTCTCGTTCCTACTCCATCAACTGATTATG
 TCCAAGTTTCTCATTTATTTTAAAGCTCTCATTACGTGTTGATCATGTAATTTGAGGAT
 TGATTAAATTGAGTCTTGTGGTTTTGTGAGGACTCACAATCTTTCTATTAAAAA
 AAAAAAAAAA

>G577 Amino Acid Sequence (domain in AA coordinates: TBD)
 MGCTASKLDSSEDAVRRCKERRRLMKDAVYARHHLAAHSDYCRSLRLTGSALSSFAAGEP
 LSVSENTPAVFLRPSSQDAPRVPSSSHSPEPPPPPIRSKPKPTRPRRLPHILSDSSPSSS
 PATSFYPTAHQNSTYSRSPSQASSVWNWENFYPPSPDSEYFERKARQNHKRPSPDYDA
 ETERSDDHYCHSRDDAAEEVHCSEWGDDHDFRFTATSSSDGDGEVETHVSRSGIEEPEVKQ
 PHQDPNGKEHSDHVTSSDCYKTKLVVRHKNLKEILDAVQDYFDKAASAGDQVSAMLEIG
 RAELDRSFSKLRKTVYHSSSVFNSLSASWTSKPPLAVKYKLDASTLNDEQGLKSLCSTL
 DRLLAWKKLYEDVKAREGVKIEHEKKLSALQSQYKGGDESKLDKTKTSITRLQSLIIV
 SSEAVLTTSNAILRLRDTDLVPQLVELCHGLMYMWSMHEYHEIQNNIVQQVRGLINQTE
 RGESTSEVHRQVTRDLESASVSLWHSSFCRIIKFQREFICSLHAWFKLSLVPLSNGDPKKQ
 RPDSEFALCEEWKQSLERVPDTPVASEAIKSFVNVVHVISIKQAEVVKMKRTESAGKELEK
 KASSLRSIERKYQAYSTVGIGPGEVLDSDPLSEKKCELAACQQRQVEDEVMRHVKA
 VTRAMTLNQLTGLPNVQALTSFSSSLFTESLQTVCSRSYSIN*

>G674 (1..786)-

ATGGTGTTTAAATCAGAAAAATCAAACCGGGAATGAAATCAAAGGAGAAGCAAAGGAAG
 GGATTATGGTCACCCGAGGAAGATGAGAAGCTTAGGAGTCATGTCCTCAAATATGGCCAT
 GGATGCTGGAGTACTATTCTCTCAAGCTGGATTGCAGAGGAATGGGAAGAGTTGTAGA
 TTAAGGTGGGTTAATTATTTAAGACCTGGACTTAAGAAGTCTTTATTCACTAAACAAGAG
 GAACTATACTTCTTTCACTTCATTCCATGTTGGGTAACAAATGGTCTCAGATATCGAAA
 TTCTTACCAGGAAGAACCGACAACGAGATCAAAAATATTGGCATTCTAATCTAAAGAAG
 GGTGTAACCTTTGAAACAACATGAAACCAAAAAACATCAACACCTTTAATCACAAC
 TCACTTGAGGCCCTGCAGAGTTCAACTGAAAGATCTTCTCATCTATCAATGTGCGAGAA
 ACGTCTAATGCTCAACCTCAAGCTTTTCGCCAATCTCGTGTCTCGGAATGGTTAGAT

CATAGTTTGCTTATGGATCAGTCACCTCAAAAGTCTAGCTATGTTCAAAATCTTGTTTTC
CCGGAAGAGAGAGGATTCAATTGGACCATGTGGCCCTCGTTATTTGGGAAACGACTCTTIG
CCTGATTTCTGCGCAAATTCAGAAATTTTGTGGATGATGAGATATCATCTGAGATCGAG
TTCTGTACTTCATTTTCAGACAACTTTTTGTTCGATGGTCTCATCAACGAGCTACGACCA
ATGTAA

>G674 Amino Acid Sequence (domain in AA coordinates: 20-120)
MVFKSEKSNREMKSKKEQRKGLWSPEEDEKLRSVHLKYHGHCWSTIPLQAGLQRNGKSCR
LRWVNYLRPGLKKSFLTQKEETILLSLHSLGNKWSQISKFLPGRDNEIKNYWHSNLKK
GVTLKQHETTKKHQTPLITNSLEALQSSTERSSSSINVGETSNAQTSSFSPNLVFSEWLD
HSLMDQSPQKSSYVQNLVLPPEERGFIPGCPRYLGNDSLPDFVPNSEFLDDEISSEIE
FCTSFSDNFLFDGLINELRPM*

>G736 (1..513)
ATGGCGACTCAAGATTCTCAAGGGATTAAACTCTTTGGCAAACTATTGCATTTAACACT
CGAACAAATAAAAAATGAAGAAGAGACACACCCGCCGAGCAAGAAGCCACAATAGCCGTT
AGATCATCATCATCATCGGATCTGACGGCCGAGAAGCGTCCGATAAGATCATAGCATGT
CCAAGATGCAAGAGCATGGAGACAAAGTTCTGTTACTTCAACAACACGGAATCAG
CCTCGACACTTTTGTAAAGGCTGCCACCGTTACTGGACCGCCGGTGGTGCCTCCGGAAC
GTTCCCGTCCGGCCGGTCTCGTCCGAAGTCCAAACCACCTGGTCTGTCGTTGGTATG
CTTGGAGATGGAAATGGTGTTCGCCAAGTCGAGCTTATAAATGGCTTGCTCGTTGAGGAG
TGGCAGCATGCCGACCGCAGCTCACGGTAGTTTCCGGCATGATTTTCCCATGAAGCGG
CTCCGGTGTTACTCCGACGGTCAATCGTGTGA

>G736 Amino Acid Sequence (domain in AA coordinates: 54-111)
MATQDSQGIKLFKGTIAFNTRTIKNEETHPPEQEATIAVRSSSSDLTAEKRPDKIIAC
PRCKSMETKFCYFNNGNQPRHFCCKGCHRYWTAGGALRNVVPGAGRRKSKPPGRVVVGM
LGDGNGVRQVELINGLLVEEWQHAAAAAHGSFRHDFPMKRLRCYSDGQSC*

>G903 (96..1496)
CCCGGGTCCGACCGCGTCTCTCTCTGAACTATACAAAAACCTACTTTTAAT
TTCTCTTCCAAGAGTCAAGAACCCAGAAAGACATGACAAGTGAAGTTCTTCAAACAA
TCTCAAGTGGATCAGGTTTTGCTCAGCCACAGAGCTCATCAACCCTGGATCATGATGAAT
CTCTCATCAATCCTCTCTTGTAAAGAAAAGAGAAATCTCCCTGGAAATCTGATCCGG
AAGCTGAAGTGATAGCTTTATCCCCACGACCTTGATGGCTACGAACCGGTTCTATGTG
AGGTATGTGGCAAAGGTTTCCAAAGAGACCAAACTTACAGCTTCATCGGCGAGGACATA
ATCTTCCATGGAAGTTGAAGCAGAGGACAAGCAAAGAAGTGAGAAAACGTGTCTACGTTT
GCCCCGAGAAGACATGTGTCCACCATCACTCCTCTAGAGCTCTAGGCGATCTCACTGGAA
TCAAAAAGCATTTTTGCCGGAAACACGGGGAGAAGAAGTGGACGTGCGAGAAATGTGCTA
AGAGATACGCAGTCCAATCTGATTGGAAAGCTCATTCCAAGACTTGTGGTACTAGAGAGT
ACCGTTGCGATTGTGGCACCATTCTCAAGGCGAGACAGCTTTATCACTCATAGAGCTT
TCTGCGATGCCTTAGCGGAAGAAACCGCTAAGATAAACGCAGTGTCTCATCTCAACGGTT
TAGCCCGCGGCTGGAGCCCCAGGATCAGTTAATCTCAACTATCAATATCTCATGGGAACAT
TCATCCCACCGCTTCAACCATTGTACCACAACCGCAAACAAATCCAAACCATCATCATC
AACATTTTCAGCCACCAACTTCTCGTCGCTCTCTATGGATGGGACAAGATATCGCGC
CGCTCAACCGCAACCGGACTACGATTGGGTTTTTGGAAACGCTAAGGCAGCGTCTGCTT
GCATTGATAATAATAACTACGATGAGCAGATTACGCAAAACGCAACGCAAGTTTGA
CCACTACCACTACTCTCTGCCCCCTCTTTATTCAGCAGCGACCAACCACAAAACGCAA
ACGCAAAATCAACCGTGAATATGTCCGCGACAGCTTACTACAGAAAGCTGCTGAAATG
GCGCTACTTCTACAACAACCGCAGCGACCAATGACCCATCAACGTTTCTTCAAAGTTTCC
CGCTTAAATCCACCGATCAAACACCAGTTATGACAGTGGCGAAAAGTTTTTGTCTTGT
TCGGGTCTAACAACAACATTGGGTTAATGAGTCGTAGTCATGATCATCAAGAGATCGAGA
ACGCTAGAAATGACGTTACGGTTGCGTCTGCTTGGATGAATTACAGAATTACCCCTTGA
AACGTAGAAGAGTTGATGGTGGAGGTGAAGTGGGTGGAGGAGGGCAAACTCGGGATTTC
TCGGGGTTGGGTGTACAAACGTTGTGCCATCCATCGTCTATCAATGGATGGATTGAAAGA
GTTTAAATTTTCGGGTTAATGCATAAATTACGTAAAGAAGAAGGAATCTTTTGTCTAT
TTCCACCATTTTCTAAGATAACATATGTATATGGTAATGGAAGTTGTTTTCTTTATTAA
TTCAATATTCTAAACTTATGATATATGTATAATGAATGTGTTTATCTTCAA

>G903 Amino Acid Sequence (domain in AA coordinates: 68-92)
MTSEVLQTISSSGSFGAQQSSSTLDHDESLINPPLVKKRNLPGNPDPEAEVIALSPPTL
MATNRFCEVCGKGFQDQNLQLHRRGHNLPWKLKQRTSKEVRKRVYVCPEKTCVHHSS

RALGDLTGIIKKHFCRKHGEKKWTCEKCAKRYAVQSDWKAHSKTCGTREYRCDCGTIFSR
 DSFITHRAFCDALAEETAKINAVSHLNGLAAGAPGSVNLNYQYLMGTFIPLQPFVPPQ
 QTNPNHHQHFPPTSSSLSLWMGQDIAPPQPQPDYDWVFGNAKAASACIDNNNTHDEQI
 TQNNANASLTNTTSLAPSLSFSSDQPNANANSNVNMSATALLQKAAEIGATSTTTAATND
 PSTFLQSFPLKSTDQTTSYDSGEKFFALFGSNNNIGLMSRSHDHQEIENARNVDVTASAL
 DELQNPWKRRRVDGGGEVGGGGQTRDFLGVGVQTLCHPSSINGWI*

>G917 (32..679)

TTAGGGTTTTAGAAAGATAGATCGATTGAAGATGAGGAAAGGTAAGAGAGTGATAAAAAA
 GATAGAGGAGAAAAATAAAGAGACAAGTGACATTCGCAAAGAGAAAGAGTCTAATCAA
 GAAGGCATATGAACCTCTCTGTTCTCTGCGATGCCACCTTGGTCTCATCATCTTCTCTCA
 CTCCAACAGGCTCTACGATTTCTGCTCCAACCTCTACCAGCATGGAGAATCTCATCATGAG
 ATACCAAAAGGAAAAAGAAGGTCAAACCACTGCAGAACACAGTTTCCACTCGGATCAGTG
 TTCAGATTGCGTGAAGACGAAGGAATCAATGATGAGAGAGATAGAGAATCTTAAGCTGAA
 TCTTCAATTGTACGACGGACATGGCTTGAATCTCTTGACCTACGACGAGCTCCTTTCTTT
 TGAGCTCCATCTCGAATCTTCTCTACAACATGCTCGAGCTCGCAAGTCTGAGTTCATGCA
 TCAGCAGCAGCAGCAACAAGAGATCAAAAGCTTAAGGGAAGAAAGGGTCAAGGAAG
 CTCTTGGGAGCAGCTGATGTGGCAAGCAGAGAGACAGATGATGACGTGTCAAAGACAAAA
 AGATCCTGCGCCGGCGAATGAAGGAGGAGTTCCTTTTTTACGGTGGGGAACAACCCACCG
 ACGTCTTCCACCTCCTTAAGCTACCACAACCAGGCCCAAATACAGGCCCATAACTTCTCT
 CTATCTATAAAAAACAACCTGATAGTAAAAAGTATTGACCCGGTTTGGTTCGGTTATGTTG
 ATACCAGACTATTAATTAACCTTCGGTTAGACGTATTTACGACTTGATGCTATCTAGACCT
 TTTTGCCCTTCAAAAAA

>G917 Amino Acid Sequence (conserved domain in AA coordinates: 2-57)

MRKGKRVIIKKIEEKIKRQVTFARKKSLIKKAYELSVLCDVHLGLIIFSHSNRLYDFCSN
 STSMENLIMRYQKEKEGQTTAEHSFHSDQCSDCVKTKESSMREIENLKLNLQLYDGHGLN
 LLTYDELLSFEHLHLESSLQHARARKSEFMHQQQQQQTDQKLKGKEKGQSSWEQLMWQAE
 RQMMTCQRQKDPAPANEGGVPLRWGTTTHRSSPP*

>G921 (116..1024)

CCAAGATCGACTCTTACTTCAATCTCTCTCAACTTTCTTCTCAGCTTACGGGAACCTC
 CACACATATACATCCACAAGAACCCATATCGAAGATTCTACATATATTTACATGGA
 TCAGTACTCATCCTCTTTGGTTCGATACTTCATTAGATCTCACTATTGGCGTTACTCGTAT
 GCGAGTTGAAGAAGATCCACCGACAAGTGCTTTGGTGGGAAGAATTAAACCGAGTTAGTGC
 TGAGAACAAGAAGCTCTCGGAGATGCTAACTTTGATGTGTGACAACCTACAACGTCTTGAG
 GAAGCAACTTATGGAATATGTTAACAAGAGCAACATAACCGAGAGGGATCAAATCAGCCC
 TCCCAAGAAACGCAATCCCCGGCGAGAGAGGACGCATTTCAGCTGCGCGGTTATTGGCGG
 AGTGTGCGGAGTAGCTCAACGGATCAAGATGAGTATTTGTGTAAGAAGCAGAGAGAAGA
 GACTGTCTGTGAAGGAGAAAGTCTCAAGGGTCTATTACAAGACCGAAGCTTCTGACACTAC
 CCTCGTTGTGAAAGATGGGTATCAATGGAGGAAATATGGACAGAAAGTGACTAGAGACAA
 TCCATCTCCAAGAGCTTACTTCAAAATGTGCTTGTGCTCCAAGCTGTTCTGTCAAAAAGAA
 GGTTCAGAGAAGTGTGGAGGATCAGTCCGTGTTAGTTGCAACTTATGAGGGTGAACACAA
 CCATCCAATGCCATCGCAGATCGATTCAAACAATGGCTTAAACCGCCACATCTCTCATGG
 TGGTTCAGCTTCAACACCCGTTGCAGCAACAGAGAAGTAGCTTGACTGTGCCGGTGAC
 TACCGTAGATATGATTGAATCGAAGAAAGTGACGAGCCCAACGTCAAGAATCGATTTTCC
 CCAAGTTTCAGAAACTTTTGGTGGAGCAAATGGCTTCTTCTTAACCAAAGATCCTAACTT
 TACAGCAGCTTTAGCAGCAGCTGTTACCGGAAAATTGTATCAACAGAATCATACCGAGAA
 ATAGTTTAGCTTCAAATCCGTTAGAGTTTTTAGATTTGAATTTGTCATGAGTAAGAGAA
 AGAGAGTAGATTATAATCCNTTGTGATACTGAAAAAAAAAAAAAAAAAAAAA

>G921 Amino Acid Sequence (domain in AA coordinates: 146-203)

MDQYSSSLVDTSLDLTIGVTRMRVEEDPPTSALVEELNVRVSAENKLLSEMLTLMCDNYNV
 LRKQLMEYVNKNITERDQISPPKKRKSAREDAFSCAVIGGVSESSSTDQDEYLCKKQR
 EETVVKEKVSRYVYKTEASDTLVVKDGYQWRKYQKVTDRNPSPRAYFKCACAPSCSVK
 KKVQRSVEDQSVLVATYEGEHNHPMPSQIDSNNGLNHRHISHGGSASTPVAANRRSSLTVP
 VTTVDMIESKKVTSPTSRIDFPQVQKLLVEQMASSLTKDPNFTAALAAAVTGKLYQQNHT
 EK*

>G922 (1..1449)

ATGGTGGCTATGTTTCAAGAAGATAATGGAACATCTTCTGTAGCTTCATCACCACCTTCAA
 GTCTTCTCAACTATGTCACTCAACAGACCGACTCTCCTCGCTTCTTCATCTCCGTTTCAT

TGTCTCAAAGATCTCAAACCAGAGGAGCGTGGTCTCTACTTAATCCACCTCTTGCTAACT
 TGTGCCAACACGCTGGCTTCAGGTAGCCTCCAAAACGCTAACGCAGCGCTCGAGCAGCTC
 TCTCACCTCGCTTCTCCTGACGGCGACACGATGCAGCGAATCGCTGCTTACTTCACCGAA
 GCGCTTGCTAAACAGAATCCTTAAGTCTTGCCCTGGTCTTTACAAGGCTCTTAACGCAACT
 CAGACAAGAACTAAACAATGTCTCTGAGGAGATTATGTTAGAAGACTCTTCTTTGAGATG
 TTCCCGATACTCAAAGTCTCTTACTTGCTCACTAATCGAGCTATACTCGAGGCTATGGAA
 GGAGAGAAGATGGTTTCATGTGATTGATCTCGATGCTTCTGAGCCAGCTCAATGGCTTGCT
 TTGCTTCAAGCTTTTAACTCTAGGCCTGAAGGTCCACCTCATTTGAGAATCACTGGTGT
 CATCACCAAGGAAGTGCTTGAACAAATGGCTCATAGACTCATTGAGGAAGCAGAGAAA
 CTCGATATCCCGTTTCAGTTTAAATCCCGTTGTGAGTAGGTTAGACTGTTTAAATGTAGAA
 CAGTTGCGGGTTAAAACAGGAGAGGCCCTTAGCCGTTAGCTCGGTTCTTCAATTGCATACC
 TTCTTGCCCTCTGATGATGATCTCATGAGAAAGAACTGCGCTTTACGGTTTCAGAACAAAC
 CCTAGTGGAGTTGACTTGCAGAGAGTTCTAATGATGAGCCATGGCTCTGCAGCTGAGGCA
 CGTGAGAATGATATGAGTAACAACAATGGGTATAGCCCTAGCGGTGACTCGGCCCTCATCT
 TTGCCCTTACCAAGTTACGGAAGGACTGATAGCTTCTCAATGCTATTTGGGGTTTGTCT
 CCAAAGGTCATGGTGGTCACTGAGCAAGACTCAGACCACAACGGCTCCACACTAATGGAG
 AGGCTATTAGAATCACTTTACACCTACGCAGCATTGTTTGATTGCTTGGAAACAAAAGTT
 CCAAGAACGTCTCAAGATAGGATCAAAGTGGAGAAGATGCTCTTCGGGGAGGAGATCAAG
 AACATCATATCCTGCGAGGGATTGAGAGAAGAGAAAGACACGAGAAGCTTGAGAAATGG
 AGCCAGAGGATCGATTGGCTGGTTTGGGAATGTTCTCTTAGCTATTATGCGATGTTG
 CAGGCTAGGAGATTGCTTCAAGGGTGCGGTTTGTATGGGTATAGAATCAAGGAAGAGAGC
 GGGTGGCGAGTAATTTGCTGGCAAGATCGACCTCTATACTCGGTATCAGCTTGGAGATGC
 AGGAAGTGA

>G922 Amino Acid Sequence (conserved domain in AA coordinates:225-242)

MVAMFQEDNGTSSVASPLQVFSTMSLNRPTLLASSSPFHCLKDLKPEERGLYLIIHLLLT
 CANHVASGSLQANANAALQLSHLASPDGDTMORIAAYFTEALANRILKSWPGLYKALNAT
 QTRTNVSEIEIHVRRLPFEMFPILKVSYLLTNRAILEAMEGEKMHVIDLDASEPAQWLA
 LLQAFNSRPEGPPHLRITGVHHQKEVLEQMAHRLIEEAEKLDIPFQFNPVVSRLDCLNVE
 QLRVKTGEALAVSSVLQLHTFLASDDDLMRKNCALRFQNNPSGVDLQRVLMMSHGSAAEA
 RENDMSNNNGYSPSGDSASSLPLPSSGRTDSFLNAIWGLSPKVMVTEQSDHNGSTLME
 RLLESLYTYAALFDCLETKVPRTSQDRIKVEKMLFGEEIKNIIISCEGFERRERHEKLEKW
 SQRIDLAGFNVPLSYAMLQARRLLQCGFDGYRIKEESGCAVICWQDRPLYSVSAWRC
 RK*

>G932 (206..1213)

CCACGCGTCCGACCACCTTGACCTCTTTGTCTTAAGTACTCTTTAACCCCTACAATTTCTCT
 AAGCTCTCAAGCCACAAAAACCACAAACCGTTCTTACCAATATATATATCTGATCATC
 ATCAAAGTCTTCTCTCTGCTCATACCACAAACCGTTCCATTCTTCCCTAATCACAAAG
 TGATATTTACATAGAGAAGATAGAGATGGGAAGACCACCATGCTGTGACAAGATTGGAGT
 GAAGAAAGGACCATGGACACCAGAGGAAGATATCATCTTGGTTTCTTACATCCAAGAACA
 TGGTCTGGAAGTGGAGATCTGTGCCTACTCACACAGGTTTGAGGAGATGTAGCAAAAG
 CTGTAGATTGAGGTGGACTAATTATCTTCGACCTGGGATCAAGCGTGGAAATTTACCGA
 GCATGAAGAGAAGATGATTCTCCATCTTCAAGCTCTTTTGGGAACAGGTGGGCAGCTAT
 AGCATCATATCTTCCAGAAAGGACAGACAATGATATAAAGAACTATTGGAACACTCATTT
 GAAGAAAAAGCTCAAGAAGATGAATGATTCTTGTGATAGTACTATCAACAATGGCCTTGA
 TAATAAAGACTTCTCCATATCAAACAAAAACACTACCTCACATCAAAGCAGCAACTCCAG
 TAAAGGTCAATGGGAGAGAAGGCTTCAGACAGATATCAACATGGCTAAACAAGCTCTTTG
 TGATGCCTTGCTATTGACAAACCACAAAACCCAACTAATTTTCTATTCCCGATCTTGG
 TTATGGTCCATCATCTTCTCGTCCTCTACCACCACCACCACCACCACCACCACCAGAG
 AAACACTAATCCATACCCATCTGGGGTCTATGCTTCAAGTGTGAGAACATTGCTCGTTT
 GCTTCAGAATTTTATGAAAGACACACCAAAGACCTCGGTGCCCTTGCCGGTTGCAGCCAC
 CGAGATGGCTATCACACGGCAGCTTCGAGCCCTAGCACAAACGAAGGAGACGGAGAAGG
 GATTGACCATTCTTTGTTTCAGCTTCAACTCCATAGATGAAGCTGAAGAGAAGCCTAAACT
 AATAGACCATGACATTAATGGTCTAATTACACAAGGCTCTCTTTCTTTGTTTCGAGAAATG
 GCTCTTTGATGAGCAAGCCACGATATGATCATCAATAACATGTCACTAGAGGGTCAGGA
 AGTGTGTTCTAGAAAGCATTAAAGTTTGACGATTGCTTGGGAACCACGAGGCTTAGT
 TATAACAATTTGTATAATTAAAGTACTCTTTAGTTTGTGTTTCAATCTTATTATGATCA
 TATTGCAGTAATTAGGGATTTTAGTCTTTAGTAGTAACCTCTTAAGTTTTAACACATTTTT

CTCTATCTTTTTAGTAGTAACCTCTTTATTTTTTCCTTAAATCTTTGTGCGACGTGGAGATG
ATATCTTCTATGTAGTAGAACTCAAAAGTGACATCATCTTTATTAATGTAACGTCTTT
TTAAAAAAAAAAAAAAAAAAAA

>G932 Amino Acid Sequence (domain in AA coordinates: 12-118)
MGRPPCCDKIGVKKGPWTPPEEDIILVSYIQEHGPGNWRVPTHTGLRRCSKSCRLRWNTNY
LRPGIKRGNFTEHEEKMILHLQALLGNRWAAIASYLPERTDNDIKNYWNTHLKKLKKMN
DSCDSTINNGLDNKFISISNKNVTSHQSSNSSKGQWERRLQTDINMAKQALCDALSIDKP
QNPTNFSIPDLGYGPSSSSSSSTTTTTTTTTTRNTNPYPSPGVYASSAENIARLLQNFMDT
PKTSVPLPVAATEMAITTAASSPSTTEGDGEGIDHSLFNFNSIDEAEKPKLIDHDINGL
ITQGSLSLFEKWLFDQSHDMIINNMSLEGQEVLF*

>G599 (152..1579)

TCGACAGAACAGCTTCGTTGTCACCTTGTCTATTCTATAAATCGCATCCCCATTGACAACCT
TTCACCTTCCATCAAACTCTCTCTCTATATCTCTCTCTATATCTCTCTCTATATCT
CTCTCTCTCTTCACTCTCTCTTTCTTTCAAATGGAAAACTCATGGTCCGACATGGAG
ACCCGACCCGGTTTACCGTCCACCGGAAACACCACTCGAACCGATGGAGTTTTAGCTCG
TTCATGGAGCGTCTCTGCTCTCGAAGTCTCAAGGCTCTAACACCACCAACCTCAGAT
TCTCTCTCTCAAAACCGAAGAAGAAGAAGAAGAAGAACCCATCTCCTCTGTCTAGACGG
CGACGGCGACACGGAAGACACCGGACTTGTACCGGAAACCCATTCTCCTTCGCTTGTTC
AGAAACTTCTCAAATGGTCATGGATCGTATCTTGTCTCACTCTCAAGAAGTATCACCAG
AACATCTGGTCGGCTATCTCACAGTAGTGGTCCACTTAATGGTTCTTTGACCGACAGTCC
TCTGTGTCTCTCCCGAATCCGACGACATTAAGCAATTTTGCAGAGCGAACAATAATTC
ATTGAACAGTGTAAATCTCAGTTCGGTTCACGGCGGCAACTCCGGGACCTATAACCGC
TACAGCTACACAGTCCAAGACGGTGGGACGGTGGCTTAAGGACCGGAGAGAGAAAAAGAA
AGAGGAGACTCGGGCTCATAACGCTCAGATTACGCTGTCTGTCTGTCTCGCGCGCTTGC
TGCAGCTGTTGCTGCTATTGTCAGCAGCCACCGCTGCGTCTTCTAGCTGTGGTAAGGATGA
GCAGATGGCTAAAACTGACATGGCCGTTGCTTCTGCTGCGACCCTTGTGGCTGCTCAGTG
TGTGGAAGCTGCTGAAGTTATGGGAGCTGAGAGAGAGTATTGGCTTCTGTTGTTAGCTC
CGCCGTCAATGTTCTGCTTCTGCCGGAGATATTATGACTCTCACCGCCGGAGCAGCTACAGC
TTTAAGAGGAGTGCAAAACATTGAAGGCAAGGGCAATGAAGGAAGTGTGGAACATAGCATC
AGTGATACCAATGGATGAAGGACTCACTTCTACAGGAGGAAGCAGCAATAATGTTAATGG
TAGCAATGGAAGCTCAAGCAGTAGTCAAGTGGTGAACCTGTACAAACAGGAGAATTTCTT
GGGAACCTGTAGTAGAGAATGGCTCGCTAGAGGTTGTGAACCTCTCAAACGCACTCGCAA
AGGTGATCTCCACTGGAAGATAGTATCTGTTTACATCAACAAAATGAATCAGGTTATGTT
GAAGATGAAGAGCAGGCATGTTGGAGGAACCTTCACCAAGAAGAAAAAGAACATTGTGCT
TGATGTGATCAAGAATGTCCCGGCTGGCCTGGACGACATTGCTAGAGGGAGGAGATGA
TCTAAGATACTTCGGTTTGAAGACGGTTATGCGAGGTGATGTTGAATTCGAGGTCAAGAG
CCAAAGGGAATATGAATGTGGACACAAGGTGTCTCAAGGCTCTTGTCTTGTCTGCTGCTGA
GAGGAAGTTTAGGATGTGAATAAACGTTCAATGGCTGCTTGGTTAAGTGTGAGTTTTTTT
TTTAACTTATGTGGTCAAATTTCAATAGTAGGGGTTCTTTTAAAGGTAATGGTTTTTTGGG
TTGGGTATAGGATAAAATGGACCTACAGTCAAGGTGAGGAAGCATTGGGTAAACAAAA
CTTAGTGGGGGTGATCTGTAATATCTATGTTCTTAGTTTTTTTTTGGTTGTGGTGGTCT
TTTTGTATAAAAAAACAAAGTTGAAGTAATAGATATATAGTATGTTTTTAATTTTAA

>G599 Amino Acid Sequence (domain in AA coordinates: 187-219, 264-300)

MEKLMVPTWRPDPVYRPPETPLEPMEFLARWSVSALEVSKALTPPNPQILLSKTEEEEE
EPISSVVDGDDTETGLVTGNPFSSACSETSQMVMDRILSHSQEVSPRTSGRLSHSSG
PLNGSLTDSPPVSPPESDDIKQFCRANKNSLNSVNSQFRSTAATPGPITATATQSKTVGR
WLKDRREKKKEETRAHNAQIHAASVAVGVAAGVAAIAAATAASSSCGKDEQMAKTDMAVA
SAATLVAAQCVEAAEVMGAEREYLASVVSSAVNVRSAGDIMTLTAGAATALRGVQTLKAR
AMKEVWNIAVIPMDKGLTSTGGSSNNVNGSNGSSSSSHSGELVQQENFLGTCSEWLAR
GCELLKRTKRGDLHWKIVSVYINKMNQVMLKMKSRHVGGFTFKKKKNIVLDVIKNVPAPW
GRHLLLEGGDDLRYFGLKTVMRGDVEFEVKSQREYEMWTQGVSRLLVLAAERKFRM*

>G804 (114..1139)

ATACTCCAAGAAATTTATAGGTTATAAGTAAAAATTCAGTACAAGTTTGTGTTGTTT
TTCCATTTTCTGTGTGTTTTTTTCCCATTAATTTATAAATTTTATAAGCAATATGGAGT
CCCACAACAACACAGAGCAACAACAACACCACTGGTTCGGCCCATCTGGTCCCATCCA
TGGGACCAATCTCCGGTTCAGTCTCATTAAACCACTGCTCCAACTCCACTACCACCA
CCGTACACCGCCGCTAAACACCCGCAAAACGACCGTCCAAGGACCGTCACATCAAAGTAG

ACGGACGTGGCCGGAGGATACGTATGCCGGCTATCTGCGCAGCACGTGTCTTCCAACATA
CACGTGAGTTACAACACAAATCGGACGGCGAGACTATAGAGTGGCTGCTCCAACAAGCGG
AGCCAGCTATCATCGCAGCCACCGGAACCTGGAACCATACCGGCGAATATCTCTACTTTGA
ACATCTCTCTTCGAAGCAGTGGCTCTACTCTTTCAGCTCCACTGTCTAAATCTTTCCACA
TGGGAAGAGCGGCTCAAAACGCTGCCGTTTTTGGGTTCCAGCAACAGCTTTATCATCCTC
ATCATATCACGACAGATTCTTCTTCTTCTCTTCCCAAACATTCCGTGAAGAAGATC
TTTTTAAAGATCCTAATTTTCTAGATCAAGAACCCGGTTCAAGATCACCTAAACCGGGAT
CCGAAGCTCCTGATCAAGATCCGGGTTCCGACCCGGTCAAGAACACAAATATGATACCGC
CGATGTGGGCACTAGCGCCAACGCCAGCCTCCACAAACGGAGGTAGTGCTTTTGGATGT
TACCAGTCGGAGGAGGAGGAGGTCCGGCTAACGTTTCAAGATCCATCACAGCACATGTGGG
CGTTTAATCCGGGTCATTACCCGGGTCGAATCCGGGTCGGTTTCAAGTGGGTTCTATGTTAG
TGGGAGGTCAACAGTTAGGGTTAGGTGTTGCAGAAAATAACAATTTGGGGCTATTTTCCG
GCGGAGGAGGACGGTTCGGGTTGGTCTCGGAATGAGTCTTGAGCAAAAGCCTCAAC
ATCAAGTGAGTGATGCTACTAGAGACCAAAATCCTACTATAGATGGTTCTCCTTGAA
AGACTTCATGATTTCTTTGGTTTTTAAAAAGTGTGAATGTGTGATTTATTGCAACTTTTG
TTGAGGACTCCAATGTTAATATGGGTTTTAGGGTTGGCTTTTTCGGGATTGCCAATTGTT
ATT

>G804 Amino Acid Sequence (domain in AA coordinates: 54-117)
MESHNNQSNNTTGSALHVPMSGPISGSVSLTTAPNSTTTTVAAKTPAKRPSKDRHI
KVDGRGRRIRMPAICARVFLQTLRELQHKSDGETIEWLLQQAEPALIAATGTGTIPANIS
TLNISLRSSGSTLSAPLSKSFHMGRAAQNAAVFGFQQQLYHPHHITDSSSSSLPKTFRE
EDLFKDPNFLDQEPGSRSPKPGSEAPDQDPGSTRSRTONMIPPMWALAPTPASTNGGSF
WMLPVGSGGGPANVQDPSQHMWAFNPGHYPRIGSVQLGSMVLGGQQQLGLGVAENNNLGL
FSGGGDGGRVGLGMSLEQKPQHQVSDHATRDQNPTIDGSP*

>G1062 (297..1781)

CAAAAAAAGTTTCAATTTTGAAGCTCTGAGAAATGAAATCTATCATTCTCTCTCTC
TATCTCTATCTTCTTTCAGATTTTCGTTCTTCAATTCATGAAATCCTCGTGATTCTAC
TTTAATGCTTCTCTTTTACTTTTCCAAGTCTCTGAATATTCAAAGTATATATCTTTT
GTTTTCAAACCTTTTGCAGAATTGTCTTCAAGCTTCCAAATTTCAAGTTAAAGGTCTCAACT
TTGCAGAATTTTCTCTAAAGGTTTCAAGCTTTGGGGTAAAGGTGTCAACTTTGGCGATGG
GTCTTGACGGAAACAATGGTGGAGGGTTTGGTTAAACGGTGGTGGTGGAGAAAGGGAAG
AGAACGAGGAAGGTTTCAAGGGAAGGAATCAAGAAGATGGTCTTCTCAGTTTAAGCCTA
TGCTTGAAGGTGATTGGTTTAGTAGTAACCAACCACATCCACAAGATCTTCAGATGTTAC
AGAATCAGCCAGATTTTCAAGTACTTTGGTGGTTTTCTTTTAAACCTTAATGATAATCTTC
TTCTTCAACACTCTATTGATTCTTCTTCTTCTTGTCTCTCTTCTCAAGCTTTTAGTCTTG
ACCTTCTCAGCAAAATCAGTTCTTGTCAACTAACAACAACAAGGGTGTCTTCTCAATG
TTCTTCTTCTGCAAAACCTTTTGATAATGCTTTTGAGTTTGGCTCTGAATCTGGTTTC
TTAACCAATCCATGCTCCTATTTTCGATGGGGTTTGGTTCTTTGACACAATTGGGGAA
GGGATTTGAGTTCTGTCTCTGATTTCTTGTCTGCTCGGTCACTTCTTGCCTCGGAAAGCA
ACAACAACAACAATGTTGTGTGTGGTTTTCACAGCTCCGTTGGAGTTGGAAGGTTTGG
GTAGTCTCTGATAATGGTGGTTTTGTTGGGAACAGAGCGAAAGTTCTGAAGCCTTTAGAGG
TGTTAGCATCGTCTGGTGACAGCCTACTCTGTTCCAGAAACGTGCAGCTATGCGTCAGA
GCTCTGGAAGCAAAATGGGAAATTCGGAGAGTTCCGGAATGAGGAGGTTTAGTGATGATG
GAGATATGGATGAGACTGGGATTGAGGTTTCTGGGTTGAATATGATCTGATGAGATAA
ATGAGAGCGGTAAAGCGGCTGAGAGTGTTCAGATTGGAGGAGGAGGAAGGGAAGGAAGA
AAGGTATGCCTGCTAAGAATCTGATGGCTGAGAGGAGAAGGAGGAAGAAGCTTAATGATA
GGCTTTATATGCTTAGATCAGTTGTCCCCAAGATCAGCAAAATGGATAGAGCATCAATAC
TTGGAGATGCAATGATTATCTGAAGGAACCTTCTACAAAGGATCAATGATCTTCACAATG
AACTTGAGTCAACTCCTCTGGATCTTTGCCCTCAACTTCATCAAGCTTCCATCCGTTGA
CACCTACACCGCAAACTCTTCTTGTCTGTCAAGGAAGAGTTGTGTCCCTCTTCTTTAC
CAAGTCTTAAAGCCAGCAAGCTAGAGTTGAGGTTAGATTAAAGGAAGGAAGAGCAGTG
ACATTATATGTTCTGTGGTCTGATACCGGCTGTTGCTGCTACCATGAAAGCTTTGG
ATAATCTTGGATTGGATGTTTCAGCAAGCTGTGATCAGCTGTTTAAATGGGTTTGCCTTGG
ATGTTTTCCGCGCTGAGCAATGCCAAGAAGGACAAGAGATACTGCCTGATCAAATCAAAG
CAGTGCTTTTTCGATACAGCAGGGTATGCTGATGATCTGATCTGATCCTGACTTCGAGT
CCATTAAGCATCTGTTGAAGCAGAGCTAGAAGAATAAGTCCCTTTAAATCTGCAATTTT
CTTCTCAACTTTTTTTCTTATGTATAACTTCAATCTAAGCATGTAATGCAATTGCAAT

>G1521 (1..891)

ATGCCTCCATTACCGTCTCCACGGCGCCTTCGTCTTCGAGACATCTTCGATCGCCGGAA
AGTATCGCGAAATTTGCAGGGAGAGCAATATTTCTGCTTTACAGGGGAAATCGTGTCG
ATATGCCTCGAAAACTAACCAGAGCGAAGATCCGCCGCCGTGATCACGGTGTGCAAGCAC
GGATACTGCCTTGCTTGATTTCGGAAGTGGAGCAGCTTCAAGAGGAATTGTCTCTTTGT
AACACTCGTTTTGATTCTGGTTTATCGTTAGTGATTTTGCTTCTAGAAAATACCATAAG
GAGCAATTACCAATTCTTCGTGATCGTGAGACTTTAACTTATCATCGGAATAATCCTTCC
GATCGCCGGAGGATAAGTTCAAAGGTGAGGGATGTTTTGGAAAACCTCTAGCTCAAGATCA
AGGCCATTGCCATGGCGGAGATCATTTGGACGACCAGGTTCACTTCTGATTCTGTTATC
TTCCAGCGAAAGCTTCAGTGGCGAGCTAGCATATACACTAAGCAATTACGAGCTGTTTCA
TTACATTCAAGGCGCTTGGAACTAAGTTTGGCGGTGAATGATTACACCAAGCAAAGATA
ACTGAAAGAATTGAGCCATGGATTAGAAGAGAGCTTCAGGCAGTCCCTTGGAGATCCTGAT
CCCTCAGTTATTGTTTCAATTTTGCCTCAGCTCTTTTCATCAAAGGCTTGAGAGAGAGAAT
AATCGACAAACCGGCGAGACCGGGATGTTGGTGGAGATGAAGTCTCCTCTCTTCGAAAA
TTCTTGCTGATAAGTTGGAATATATTTTGGCATGAACCTAAGATGTTTTGCGGAGAGTATA
CTCACGATGGAGACTTATGATGCAGTGGTTGAATACAATGAGGTGGAGTAA

>G1521 Amino Acid Sequence (domain in AA coordinates: 39-80)
MPPLPSSTAPSSSRHLRSPESIAKFAGRAIFPALQKSCPICLENLTERRSAAVITVCKH
GYCLACIRKWSFKRNCPLCNTRFDSWFIVSDFASRKYHKEQLPILRDRETLTYHRNNPS
DRRRIIQSRDVLNSSSRSLPWRRSFGRPGSVPSVIFQRKLQWRASIYTKQLRAVR
LHSRRLSLAVNDYTKAKITERIEPWIRRELQAVLGDPSVIVHFASALFIKRLEREN
NRQTGQTGMLVEDEVSSLRKLFLSDKVDIFWHELRCFAESILTMETYDAVVEYNEVE*

>G183 (1..1458)
ATGAGTGATTTTGATGAAAACCTTCATCGAAATGACGTCGTATTGGGCTCCACCATCCAGT
CCTAGCCCAAGAACGATATTGGCAATGCTGGAGCAAACCGACAATGGTCTGAATCCAATC
AGTGAGATCTTCCCTCAGAAAGCTTGCCAAGAGATCATACTGATCAATCTGGACAAAGA
TCTGGTCTTCGTGAGAGACTGGCTGCAAGAGTAGGATTCAATCTTCCAACACTCAATACA
GAAGAAAACATGAGTCTTGGATGCATTTTTCAGGAGCTCGAATGTTCTTAATCTCCT
GTCGTTGCAATCTCTCCAGGATTCAGTCCATCAGCACTATTGCATCTCCCAATATGGTC
AGTGATTCTTCCAGATTATCCCTCCGTCTTCAGCCACCAATTACGGACCTCTAGAGATG
GTGGAACCTCCGGTGAAGACAATGCAGCGATGATGATGTTCAACAACGATCTTCTTAT
CAGCCGTACAATGTTGATCTGCCTTCTCTAGAAGTCTTTGATGATATTGCAACGGAAGAG
TCCTTTTATATCCCATCTTATGAACCTCATGTTGACCAATTGGAACCTCTTTAGTCACA
TCCTTTGAATCTGAACCTGTTGACGATGCCCATACCGACATCATCTCCATTGAGGACAGT
GAGACGAGGATGGAACAAAGATGATGACGACGAGGACTTCCAATACGAAGACGAAGAC
GAAGACCAATACGACCAGATCAAGATGTAGATGAAGATGAAGAGGAAGAAAAAGATGAA
GACAATGTTGCATTAGATGATCCTCAACCTCCACCTCCAAAGAGAAGGAGATATGAGGTA
TCAAACATGATTGGAGCCACAAGAACAAGCAAGACACAAAGGATCATACTTCAGATGGAA
AGCGACGAAGACAATCCTAACGATGGTTATCGCTGGAGAAAATACGGTCAGAAAGTCGTC
AAAGGAAATCCTAATCCGAGGAGTTACTTCAAGTGACAAACATCGAGTGACAGTGAGAA
AAACATGTGGAGAGAGGAGCAGACAATATCAAGTTGGTTGTGACTACATACGATGGGATA
CACAACCATCCTTCACCACCTGCACGTAGAAGCAATTCCAGTTCAAGGAACCGGTCTGCA
GGGGCAACAATACCTCAAATCAGAATGATCGAACCAGTCGGTTAGGTAGGGCTCCTCCT
ACTCCTACTCCTCCTACTCCTCCTCCTCGTCTTACACACCTGAGGAGATGAGGCCTTTC
TCTTCGTTGGCTACAGAAATTGATCTGACAGAGGTTTATATGACCGGAATCTCTATGCTG
CCGAATATACCGGTTTACGAGAATTCGGGTTTATGTACCAGAATGATGAACCGACGATG
AATGCGATGCCGATGGTTCAGATGTGTACGATGGGATCATGGAACGCCTGTATTTAAG
TTTGGTGTGACATGTAG

>G183 Amino Acid Sequence (domain in AA coordinates: TBD)
MSDFDENFIEMTSYWAPSSPSRPTILAMLEQTDNGLNPISIEIFPQESLPRDHTDQSGQR
SGLRERLAARVGFNLPPTLNTEENMSPLDAFFRSSNVNPSVVAISPFGSPSALLHTPNMV
SDSSQIIPSSATNYGPLEMVETSGEDNAAMMMFNNDLPYQPYNVDLPSLEVFDIATEE
SFYIPSYEPHVDPITPLVTSFESELVDDAHTDIISIEDSESEDGNKDDDEDFQYEDD
EDQYDQDQVDEDEEEKEDNVALDDPQPPPKRRRYEVSNMIGATRTSKTQRIILQME
SDEDNPNNDGYRWRKYQKVVKGNPNPRSYFKCTNIECRVKKHVERGADNIKLVVTTYDGI
HNHPSPPARRSNSSSRNRSAGATIPQNQNDRTSRLGRAPPTPTPTPTPPSSYTPPEMRPF
SSLATEIDLTEVYMTGISMLPNIPVYENS GFMYQNDEPTMNAMPDGSVDYDGMERLYFK
FGVDM*

>G2555 (177..956)

CTGTTTTTGTATCCGTGTAAATTAATCACACGGTAGTTTTTGTATGAAAAGACAACAATCG
GAGAACAATCTGGTCTGCTGCTAAAAATTTAATAAATGTTTTGTCTAATTGTCTCCACCC
ATAAAAAAGCGCGAATTCATTCACCGACTAAAGACATTCTCCGGTGGAGACCCCGATGC
AATCCACTCATATAAGCGGCGGAAGTAGCGGTGGTGGTGGTGGAGGAGGAGGAGGTGA
GTCGAAGTGGATTATCTCGGATCCGTTCCAGCTCAGCTACTTGGATTGAAACCCTACTCG
AAGAAGATGAAGAAGAAGGTTTAAACCTAACCTTTGTTTAAACAGAGCTGCTTACTGGTA
ATAATAACTCTGGAGGAGTGATAACGAGTCGTGACGACTCGTTCGAGTTCCTGAGTTCCTG
TTGAGCAAGGATTGTATAATCATCATCAAGGTGGTGGCTTTCACCGTCAGAATAGTTCTC
CGGCTGATTTTCTTAGTGGGTCTGGTTCGGGACTGATGGGTATTTCTCTAATTTTGGTA
TTCCGGCGAATTATGACTATTTGTGCGACCAACGTTGATATTTCTCCGACTAAACGGTCTA
GAGATATGGAACACAGTTTTTCTTCTCAGCTGAAAGAAGAGCAAATGAGTGGTGGGATAT
CAGGAATGATGGATATGAACATGGACAAGATTTTTTGAGGATTGAGTTCCTTGTAAGGTTT
GTGCTAAACGTGGTTGTGCTACTCATCTCGTAGCATTGCTGAACGGGTGAGAAGAACGC
GAATAAGTGATCGGATTAGGAGGCTGCAAGAGCTTGTTCCTAACATGGATAAGCAAACCA
ACACTGCAGACATGTTGGAGAAGCTGTGGAGTATGTGAAGGCTCTTCAAAGCCAGATCC
AGGAATTGACAGAGCAGCAGAAGAGATGCAAATGCAAACCTAAAGAAGAACAATAATGTA
TCCTTTAGGATTTGATATATCTGTATTTTATTTTGTACTATCTAAAAATGGTGATGATC
TGTTGCAAAATTCGAAACATGATCTTATATATTGAACTAGAAAAAATAGATATATATGAA
TTTTAGCTGTAAAATTTTTGTACAATAAGGAGAGAAAAAGATTTAGAAGAGTCAATAAAAAAG
ATGATGTTTACAAGTCAAAAAA

>G2555 Amino Acid Sequence (domain in AA coordinates: 175-245)

MQSTHISGGSSGGGGGGGVSRLSRLSAPATWIELLEDEEEGLKPNLCLTELLT
GNNNSGGVITSRDDSEFLSSVEQGLYNHHQGGGFHRQNSSPADFLSGSGSGTDGYFSNF
GIPANYDYLSTNVDISPTKRSRDMETQFSSQLKEEQMSGGISGMMDMNDKIFEDSVPCR
VRAKRGCAHPRSAERVRRTIRSDIRRLQELVPMNDKQNTADMLEEAVEYVKALQSQ
IQELTEQQKRCKCKPKEEQ*

>G375 (53..1171)

TCGACAAAACTCTCACTCTCCCTCAAACATAACAAACATACAGAACACAAAATGGGTCT
CACTTCTCTTCAAGTTTGCATGGATTCTGATGGCTCCAGGAATCCGAGTCATCAGGAGG
AAGCATGTTAGACTCTTCAACGAATTCCTCCGTCAGCAGCCGACATACTAGCAGCTTGCAG
CACTAGACCACAAGCCTCGGCCGTGGCTGTAGCCGCTGCAGCTCTGATGGACGGTGGAG
GAGGCTCGCTCCACCTCAGACCATCTCAAAGTGTCTCGTTGCGAGTCAACACATAC
TAAATTTCTGTACTACAATAACTACAGCCTCTCTCAGCCTCGTTACTTCTGCAAGACTTG
TCGCCGTTACTGGACAAAAGGCGGAACCTAAGGAATATTCCGGTTGGTGGTGGATGCCG
TAAAAACAAGAAACCATCTTCTCTAATTCCTCCTCCTCACTTCTTCCGGCAAAAAACC
ATCCAACATCGTTACCGCCAATACCTCTGATCTTATGGCTTTAGCACATCTCATCAAAA
TTACCAACATTCTCCTCTAGGGTTTTACATTTTGGTGGGATGATGGGGTCTTACTCAAC
TCCGGAGCATGGTAACGTTGGTTTCTTGGAGAGCAAGTATGGCGGTTTGCTTTTCGAGAG
CCCTAGACCTATTGATTTCTTGGACAGTAAGTTTGATCTCATGGGAGTGAAATGACAA
CCTGGTCATGGTTAATCATGGAAGTAACGAGATCATCATCATCATATAATCATCACAT
GGGTCTGAATCAGGTTGTAGGTCTTAACAACAACAACAATGGTGGATTAAATGGGAT
TTCTACGGGAGGCAATGGAAATGGTGGTGGTCTCATGGATATATCGACATGCCAAAGACT
TATGCTATCTAATTATGATCATCACCATTACAATCATCAAGAAGATCATCAAGGGTAGC
AACAATAATGGATGTGAAGCCAAATCCGAAGTTGTTATCGCTTGATTGGCAGCAAGATCA
ATGCTACTCCAATGGTGGTGGTAGCGGAGGCGCAGGAAAATCCGACGGTGGTGGATACGG
CAATGGTGGTTATATCAACGGTTTAGGTTTCGTGCGTGGAAATGGTTTGATGAATGGCTATGG
AACGTCCTCAAAAAAACCCTTGGTTTGATAAGTTAATCAGAACCTCTTTTTTCTTGT
CGTCATCAACTAGTAGTAGTAATAGTAGTTGGAGACTAGAGAAGCACTTCAAATTAT
TTATGGGTTTGGTTTGCTAAGCCAGTTTTAC

>G375 Amino Acid Sequence (domain in AA coordinates: 75-103)

MGLTSLQVCMDSDWLQSESSSGSMLDSSNPSAADILAACSTRPQASAVAVAAAAALMD
GGRRLRPPHDHPQKPCRESTHTKFCYNNYSLSQPRYFCKTCRRYWTGKGTLRNIPVGG
GCRKNKPPSSNSSSTSSGKKPSNIVTANTSDLMALAHSHQNYQHSPLGFSHFGMMGS
YSTPEHGNVGFLESKYGGLLSQSPRPIDFLDSKFDLMGVNNDNLVMVNHGNSNGDHHHHHN
HHMGLNHGVLNNNNNGGFNGISTGGNGNGGGLMDISTCQRLMLSNDHHHYNHQBHDHQ
RVATIMDVKNPKLLSLDWQQDQCYSNGGSGGAGKSDGGGYNGGYINGLSSWNGLMN

63

CTTCCACACTTCCTTTTACTAGGCAGTGTTAACCAATTGAGAGAGAAAAATGATGGTTG
ATGAAAATGTGGAAACCAAGGCCTCTACTTTAGTGGCAAGTGTTGATCATGGGTTTGGAT
CCGGGTCGGGTCATGATCATCATGGGTTATCGGCGTCTGTGCCTCTCTTGGTGTAACT
GGAAGAAGAGAAGGATGCCTAGACAGAGACGATCTTCTTCTTCTTAACTTCTCTCTT
TCCCTCCTCTATGCCTCTATTTCCACGTGCCAATCCTCTCCCGCACGTAAAATTG
ACCCAGAAAGCTAAGATTCTCTTCCAAAAGGAACCAAGAACAGTGACGTGAGCTCTC
TCCGACGTATGATACTCCCGAAGAAAGCCGGGAGGCTCACTTGCCGGCACTTGAATGCA
AGGAAGGGATTCCCTATAAAGATGGAAGATTTGGACGGTTTTCACGTTTGGACCTTCAAGT
ATAGGTACTGGCCAAACAACAATAGCAGAATGTACGTGCTAGAAAACACAGGCGATTG
TGAATGCTCATGGTCTGCAGCTAGGTGACTTCATCATGGTTTACCAAGATCTCTACTCAA
ACAATTACGTTATACAAAGCAAGAAAAGCATCGGAAGAAGAAGTAGACGTAATCAATC
TTGAAGAAGACGACGTTTACACAAACTTAACAAGGATCGAAAACACTGTGGTTAACGATC
TTCTCCTCCAAGATTTTAAATCATCACAACAACAACAACAACAACAGCAACAGCAACA
GCAACAAATGTTCTTACTATTATCCAGTCATAGATGATGTCAACCAAAACACAGAGTCTT
TTGTCTACGACACGACGGCTCTTACCTCCAACGATACTCCTCTCGATTTTGGGTGGAC
ATACGACGACTACTAATAATTATTACTCCAAGTTCGGAACATTGATGGTTTGGGCTCCG
TTGAGAATATCTCTCTCGATGACTTCTACTAGATAATCAATCGATGGGCTCATGGTATC
TTGATGGTGATCAGCTATTTAATATCCTTATAATATATATAAGAATTAATGCAATTTGC
ATATATATATCAAGTGTTGTAATATAACATTACAGTTTAAAAAAAAAAAAAAAAAAAA
>G1014 Amino Acid Sequence (domain in AA coordinates: 90-172)

MVDENVETKASTLVASVDHGFSGSGHDHGLSASVPLLGVNWKRRMPQRSSSSPNL
LSFPPMPPIPIHVPTPLPARKIDPRKLRFLFQKELKNSDVSSLRMILPKKAAEAHLPAL
ECKEPIRIMEDLDGFHVWTFKYRYWPNNSRMYVLENTGDFVNAHGLQLGDFIMVYQDL
YSNNYVIQARKASEEEVDVINLEDDVYTNLRIENTTVNDLLQDFNHHNNNNNNNSN
SNSNKCSYYYPVIDDVTTNTESFVYDITLTSNDTPLDFLGHTTTTNYYSKFGTFDGL
GSVENISLDDFY*

>G1035 (103..624)

CCATAATAATATATTAAACTATATACTATAATCTTTTTACATAATAAACTTTGGGTCCT
GCGTCTTAATCATAGTACTTAATTTTCTCTGTGTGTTTTAATATGAATAATAAACTGAA
ATGGGATCTTCCACAAGTGGAATTTGCTCGTCGGTTTTCAACCACTGGTTTAGCTAACTCC
GGTTCAGAATCTGATCTCCGGCAACGTGATCTAATCGACGAGCGGAAGAGAAAGAGGAAA
CAGTCGAACAGAGAATCTGCGAGGAGTTCGAGGATGAGGAAGCAGAAGCATTGATGAT
CTCACTGCTCAGGTGACTCATCTACGTAAAGAAAACGCTCAGATCGTCGCGGAATCGCC
GTCACGACGCAGCACTACGTCACTATCGAGGCGGAGAACGACATTCTCAGAGCTCAGGTT
CTTGAACCTTAACACCGTCTCCAATCTCTTAACGAGATCGTTGATTTCTCGTGAATCTTCT
TCTTCAGGATTCGGTATGGAGACCGGTGAGGATTAATCGACGGTGGATTATTCGACGGC
GTGATGAATCCTATGAATCTAGGGTTTTATAATCAACCAATCATGGCTTCTGCTTCTACT
GCTGGTGATGTTTTCAACTGTTAGAAAACCTTCACATCATTATCATCGTGAGTGAGACTAA
TCATCGCAGCAGGGGTAAACCTGTAATTTTTCTTATAAATTATGTGATGATGCTTTGTTT
CTTTATTTTATAAGATGGTTAATTAGTGTTTAAACTGATTGTAATGATAGACAGTGTA
GAAATGTGTGATATCATGGAGATGGTGATGTGAGTTGGTACAAATATTTAAGATCTTT
TCTTTCTATATATTAAAGTGAAGAAATAATATTTTGTCAATTTCTTAAAAAAAAAAAA
AAA

>G1035 Amino Acid Sequence (domain in AA coordinates: 39-91)
MNNKTEMGSSTSGNCSSVSTTGLANSSESIDLQRDLIDERKRKRKQSNRESARRSRMRK
QKHLDDLTAQVTHLRKENAQIVAGIAVTTQHYVTIEAENDILRAQVLELNHRLQSLNEIV
DFVESSSSGFGMETGQGLFDGGLFDGVMNPMNLGFYNQPIMASASTAGDVFNCS*

>G1046 (1..567)

ATGATTAGACATCTAAAACCTACATGGAGTCGTCTAGTGTCATCGCTCTCATGTTTC
GATATTCTTGATGGAGTCCCACTACACGACGATCATTTCAACTCGGCATTCTACCAAAC
ACTGACTTTAATGTTTCATTTGCAGTCAAACGTATCGACCCGCATCAACAATCAGTCTCAC
TTAGACCCAAATGCAGAAAACATTTCCATAACGAAGGTTCTTGCTCCAGAAGAAAGAAGA
GCAAGAAGAATGGTCTCTAACCGGAATCTGCAAGGAGGTACGATATGCGCAAAAAGAAG
CAGATCGAAGAGCTGCAACAACAAGTTGAACAACCTCATGATGTTGAATCATCACTGTCT
GAGAAAGTCATCAACTTGTTGGAAAGCAACCATCAGATCCTACAAGAGAACTCACAGCTG
AAAGAGAAAGTCTCTTCTTTCACTTGCTCATGGCAGATGTGCTATTACCCATGAGAAAT
GCAGAGAGCAACATCAATGACCGCAATGTGAATTATCTAAGAGGAGAACCATCAAACCGT

CCCACCAACAGTCCCTTTGGTAAGTAA

>G1046 Amino Acid Sequence (conserved domain in AA coordinates:79-138)

MIRHLKPMESSVHRSHCFDILDGVPLHDDHFNFAFLPNTDFNVHLQSNVSTRINNQSH
LDPNAENIFHNEGLAPEERRARMVSNRESARRSRMRKKQIEELQQQVEQLMMLNHHLS
EKVINLLESNHQILQENSQLEKVVSSFHLLMADVLLPMRNAESNINDRNVNVLRGEPNSR
PTNSPFGK*

>G1049 (29..550)

CTAACTTTCTTCCCAAGTAACTTCAAAATGCAGCCGCAACAGACGTTTTTCAGCCTCCA
TAACTACCTAAACTCATCGATACTGCAGTCTCCGTATCCTTCTAATTTCCCGATATCTAC
GCCATTTCCAACCAACGGTCAAAACCCGTACCTCCTCTACGGATTCCAAAGCCCTACAAA
CAATCCACAATCCATGAGCCTAAGCAGCAACAACCTCAACATCAGATGAAGCAGAAGAGCA
GCAGACGAACAACATATAATCAACGAGCGGAAGCAGAGAAGGATGATTTCAAACCGAGA
ATCCGCAAGGAGATCGCGTATGAGGAAGCAAAGACACCTTGACGAGCTTTGGTCACAAGT
GATGTGGTTAAGGATCGAGAATCATCAGTTGCTTGATAAGCTTAACAATCTCTCTGAGTC
TCACGACAAGGTTCTTCAAGAGAATGCTCAGCTTAAAGAAGAAACATTTGAGCTTAAGCA
AGTGATCAGCGATATGCAAATTCAAAGCCCTTTCTCTTGCTTTAGAGACGATATAATCCC
CATTGAATAAAGCATTTTTTCCCGATTATATTTATGAAAATTTCTTCAAGAGTATGTT
TCTTTGTATGTATATGTTGAGATGTATTTTCAAGGTTTGTATAATGACCCCTTACGACG
ACGTTTTTAGATTGTAGTAAATTTATAAACTAAAGAAGATTAGTGTTAATGAAGAACAAA
TATAA

>G1049 Amino Acid Sequence (domain in AA coordinates 77-132)

MQPQTDFSLHNYLNSILQSPYPSNFPITPPTNGQNPYLLYGFQSPTNPNQSMSSLSS
NNSTSDEAEQQTNNNIINERKQRRMISNRESARRSRMRKQRLDELWSQVMWLRIENHQ
LLDKLNNLSESHDKVLQENAQLKEETFELKQVISDMQIQSPFSCFRDDIPIE*

>G1069 (89..934)

TTGGAACCCTAGAGGCCTTTCAAGCAAATCATCAGGGTAACAATTTCTTGATCTTTCTTT
TTAGCGAATTTCCAGTTTTTGGTCAATCATGGCAAACCCTTGGTGGACGAACCAGAGTGG
TTTAGCGGCATGGTGGACCATTCGGTCTCCTCAGGCCATCACCAAAACCATCACCACCA
AAGTCTTCTTACCAAAGGAGATCTTGGAATAGCCATGAATCAGAGCCAAGACAACGACCA
AGACGAAGAAGATGATCTAGAGAAGGAGCCGTTGAGGTGGTCAACCGTAGACCAAGAGG
TAGACCACCAGGATCCAAAAACAAACCCAAAGCTCCAATCTTTGTGACAAGAGACAGCCC
CAACGCACCTCCGTAGCCATGTCTTGAGATCTCCGACGCGAGTGACGTCGCCGACACAAT
CGCTCACTTCTCAAGACGCGAGGCAACGCGGCGTTTGCGTTCTCAGCGGGACAGGCTCAGT
CGCTAACGTCACCTCCGCCAAGCCGCGCACCAGGAGGTGTGGTCTCTCTCCAAGGCAG
GTTTGAAATCTTATCTTTAACCGGTGCTTTCTCCTCGACCTTCCCAACCCGGGTCAAC
CGGTTTAAACGGTTTACTTTAGCCGGGTCCAGGGTCAGGTGCTTGAGGTAGCGTTGTAGG
CCCCTCTTAGCCATAGGGTCGGTCATGGTGATTGCTGCTACTTTCTCTAACGCTACTTA
TGAGAGATTGCCCATGGAAGAAGAGGAAGACGGTGGCGGCTCAAGACAGATTACGGAGG
CGGTGACTCACCAGCCAGAAATCGGTAGTAACCTGCCTGATCTATCAGGGATGGCCGGGCC
AGGCTACAATATGCCGCCGATCTGATTCCAAATGGGGCTGGTCAGCTAGGGCACGAACC
ATATACATGGGTCCACGCAAGACCACCTTACTGACTCAGTGAGCCATTTCTATATATAAT
GGTCTATATAAATAAATATATAGATGAATATAAGCAAGCAATTTGAGGTAGTCTATTACA
AAGCTTTTGCTCTGGTTGGAAAAATAAATAAGTATCAAAGCTTTGTTTGTCTTAATGGA
AATATAGAGCTTGGGAAGGTAGAAAGAGACGACATT

>G1069 Amino Acid Sequence (domain in AA coordinates: 67-74)

MANPWWINQSLAGMVDHVSVSGHHQNHQSLLLTKGDLGIAMNQSQDNDQDEDDPREG
AVEVNNRRPRGRPPGSKNPKAPIFVTRDSPNALRSHVLEISDGSVDVADTIAHFSRRRQR
GVCVLSGTGSVANVFLRQAAPGGVVSLQGRFEILSLTGAFPLPGSPPGSTGLTVYLAGV
QQQVVGGSVVGPLLAIGSVMVIAATFSNATYERLPMEEEDGGGSRQIHGGDSPPRIGS
NLPDLSGMAGPGYNMPPHLIPNGAGQLGHEPYTWVHARPPY*

>G1070 (170..1144)

TCGACCAGCTTGGATTTCTGTTGTTTCATCATTACTACTCTCTTTCTTCTTCTAGCTAGCTA
GTTTTGACAGCAAAATAAGAAGCAAAAAAAGGTCAACTAAAAAAGATCTGTTCTTAGAT
CACTCTCTTCTTCTTTTGTATCCAATTCCACCATGAATCATAGATCATGGATCCAGT
ACAATCTCATGGATCACAAGCTCTTACCTCCTCCTTCCACGCAAGAGACTTCAATT
ACATCTTCAACAACAGCAACAAGAGTCTTCTCCTCACCATCACCAGCAACAAGAAACCA
AACCAGTGGTGACCAACAAGGAGGATCAGGAGGAAACCGACAAATCAAGATGGATCGTGA

AGAGACAAGCGACAACATAGACAACATAGCTAACAAACAGCGGTAGTGAAGGTAAAGACAT
AGATATACACGGTGGTTTCAGGAGAAGGAGGTGGTGGCTCCGGAGGAGATCATCAGATGAC
AAGAAGACCAAGAGGAAGACCAGCGGGATCCAAGAACAACCAAAACCACCGATTATCAT
CACACGGGACAGCGCAACCGCGCTTAGAACCCACGTGATGGAGATCGGAGATGGCTGCGA
CTTAGTCGAAAGCGTTGCCACTTTTGCACGAAGACGCCAACCGCGCTTTGCGTTATGAG
CGGTACTGGAATGTTACTAACGTCACTATACGTGAGCCTGGATCTCATCCTTCTCCTGG
CTCGGTAGTTAGTCTTACGGAAGGTTTCGAGATTCTATCTCTCTCAGGATCTTTTCTCCC
TCCTCCGGCTCCTCCTACGACCACCGATTGAGTGTTTACCTCGCTGGAGGACAAGGACA
GGTGGTTGGAGGAAGCGTAGTTGGTCCGTTGTTATGTGCTGGTCTGTGTTGTCATGGC
TGCGTCTTTTAGCAATGCGGCGTACGAAAGGTTGCCCTTTAGAGGAAGATGAGATGCAGAC
GCCGTTTCATGGCGGAGGAGGAGGAGGATCATTGGAGTCCGCCCAATGATGGGACAACA
ACTGCAACATCAGCAACAAGCTATGTCAGGTCATCAAGGGTTACCACCTAATCTTCTTGG
TTCGGTTTCAGTTGCAGCAGCAACATGATCAGTCTTATTGGTCAACGGGACGACCACGTA
TTGATCAAATATACACACACTCATAATCGTTGCTAGCTAGCTAACGATGAATCATGAG
TTTAGTGGATATATATATGATTAAAGAGGTTAGCTTATGAACATTAATAAGAGTTTGA
TTCTATCGAGCTTCATTATGTTTGGGTCATCGTTC

>G1070 Amino Acid Sequence (domain in AA coordinates: 98-120)

MDPVQSHGSQSSLPFPHARDFQLHLQQQQQEFFLHHHQQQRNQTGDGQQGSGGNRQIK
MDREETSNDINIDNIANNSSGSEKIDIDHGGSGEGGGSGGDHQMTRRPRGRPAGSKNPKP
PIIITRDSANALRTHVMEIGDGLVESVATFARRRQRGVCMVSGTGNVTNVTIRQPGSH
PSPGVSVSLHGRFEILSLSGSFLPPAPPTATGLSVYLAGGQGVVGGSVVGPPLLACAGPV
VMAASFNAAYERLPLEEDEMPTVHGGGGGSLSPMMGQQLQHQQQAMSGHQGLPP
NLLGSVQLQQQHDQSYWSTGRPPY*

>G1076 (198..1076)

ATTTTAGTCTTCTTATAAATTCTTCTCAATCCTCTCTCATATCTTTTTTCTTAGTTTAA
TTCAATAAAATAGAAAAAACATATACAAATCTACAGAGAAGAGAAGCTTTATTTTAAAT
CTTGTGTGTGTGTGTGTGTGTGTATATAATTTTATTTTTTTCAAATTAAATCTCTTCT
TTGCTTTTGTATGTTGGCATGGCTGGTCTTGATCTAGGCACAGCTTTTCGTTACGTTAATC
ACCAGCTCCATCGTCCCGATCTCCACCTTCACCACAATTCTCCTCCGATGACGTCACCTC
CCGGAGCCGGGATGGGTCAATTCACCGTCGACGACGAAGACAACAACAACCATCAAG
GTCTTGACTTAGCCTCTGGTGGAGGATCAGGAAGCTCTGGAGGAGGAGGAGTACGGCG
GGGGAGGAGACGTCGTTGGTTCGTCGTCACGTCGACGACACCACGGGATCCAAGAACAAC
CGAAACCTCCGTAATTATCACGCGGAGAGCGCAAACACTCTAAGAGCTCACATTCTTG
AAGTAACAACCGCTGCGATGTTTTCGACTGCGTTGCGACTTATGCTCGTCGGAGACAGC
GAGGGATCTGCGTTCTGAGCGGTAGCGGAACGGTCACGAACGTCAGCATACTCAGCCAT
CTGCGGCTGGAGCGGTTGTGACGCTACAAGGAACGTTTCGAGATTCTTTCTCTCTCCGGAT
CGTTTTCTCTCCTCCGGCACCTCCCGGAGCAACGAGTTTGACAATTTTCTTAGCCGGAG
GACAAGGTCAGGTGGTTGGAGGAAGCGTTTGGGTGAGCTTACGGCGGCTGGACCGGTGA
TTGTGATTGCAGCTTCTGTTTACTAATGTTGCTTATGAGAGACTTCTTTAGAAGAAGATG
AGCAGCAGCAACAGCTTGGAGGAGGATCTAACGGCGGAGGTAATTTGTTTCCGGAGGTGG
CAGCTGGAGGAGGAGGAGGACTTCCGTTCTTTAATTTACCGATGAATATGCAACCAAATG
TGCAACTTCCGGTGGAAGGTTGGCCGGGAATTCCGGTGGAAGAGGTCCTTTCTGATGTG
TATATATTGATAATCATTATATATATACCGCGGAGAAGCTTTTCCGGCGAAGAATTTGC
GAGAGTGAAGAAAGGTTAGAAAAGCTTTTAAATGACTAATGAATTTCAAATTATCATCGT
GATTTCCGACATTGTCTTGTTCATCATGTTAAGCTTAGGTTTATTTTTTGTCTTTGTAG
AATTTTATGTTGAATCCTTTTTTTTTTCTGTGAACTCTATTGTGTTCTGCTGCGAAGG
AAAAAAAATTCTCAAAAAAA

>G1076 Amino Acid Sequence (domain in AA coordinates: 82-89)

MAGLDLGTAFRYVNHQLHRPDLHLHHNSSDDVTPGAGMGHFTVDDNNDNNHQLDLAS
GGGSGSGGGGGHGGGDVVGRRPRGRPPGSKNPKPPVITRESANTLRAHILEVTNGC
DVFDCAVATYARRRQRGICVLSSGTVTNVSIHQPSAAGAVVTLQGTFEILSLSGSFLPPP
APPGATSLTIFLAGGQGVVGGSVVVELTAAGPVIVIAASFNTVAYERLPLEEDEQQQQL
GGGSGGGNLFPEVAAGGGGLPFFNLPMNQPNVQLPVEGWPGNSGGRGPF*

>G1089 (31..2427)

AAGTAAGAGAGCTTCTTAAGGAAGAAGAAGATGGGTTGTGCTCAATCAAAGATCGAGAAC
GAAGAAGCAGTTACTCGTTGCAAAGAACGAAAACAATTGATGAAAGACGCCGTCACGTCT
CGTAACGCTTTCGCCCGCGCTCACTCAGCTTACGCTATGGCTCTTAAAAACACCGGAGCT

GCTCTTTCCGATTACTCTCAGCGCGAGTTTTTAGTCTCTAATCACTCGTCTTCTCCGCA
 GCTGCAGCAATCGCTTCTACTTCTTCTCTTCCCACTGCTATATCTCCTCCTTCTTCTCT
 TCCACCGCTCCGGTTTCTAATTCAACCGCTTCTTCTTCTCCGCTGCGGTTCTCAGCCG
 ATTCTGTATCTTCTCCTCCTCCTCCTCCTCCACCACCGCTTCTCTTCAACGTGCTGCT
 ACTATGCCGGAGATGAACGGTAGATCCGGTGGTGGTCATGCTGGTAGTGACTCAACGGA
 ATTGAAGAAGATGGAGCCCTAGATAACGATGATGATGACGATGATGATGATGATGACTCT
 GAAATGGAGAATCGTGATCGTTTGATTAGGAAATCGAGAAGCCGTGGAGGTAGTACTAGA
 GGAAATAGGACGACGATTGGAAGATCATCATCTTTCAGGAGGAGAAAGCTCCGCCACCTCCC
 CCTTTGGCGAATTCGCGGCCAATTCGCCGCCACGTGACGATCAGCATCAACATCAGCAA
 CAGCAACAACAACCTTCTACGATTACTTCTTCCCTAATGTTGAGAATATGCCTGGAAC
 ACTTTAGAAGATACTCCTCCACAACCACAACCACAACCAACCAAGGCCTGTGCCTCCTCAA
 CCACATTCACCAAGTCGTTACTGAGGATGACGAAGATGAGGAGGAGGAAGAGGAGGAAGAG
 GAGGAGGAAGAGGAGACGGTGAATGACGGAACCACTGGTGGAGGAAGACCGAAGAGA
 GTAGAGGAAGTGACGATTGAATTGGAAAAAGTTACTAATTTGAGAGGGATGAAGAAGAGT
 AAAGGGATAGGGATTCCCGGAGAGAGGAGGAATGCGAATGCCGGTGAATGCGACGCAT
 TTGGCGAATGTATTCTTGAAGCTTGAAGCTTACGCTCCATTACCATTTCTAATTTGCGAGATA
 GATGTTTCTAAGATGCTTGAAGCTTACGCTCCATTACCATTTCTAATTTGCGAGATAAC
 CGAGGACATATTGATCACTCTGCTAGAGTGATGCGTGAATTACATGGAATAGATCATTT
 AGAGGAATACCAATGCTGATGATGGGAAAGATGATGTTGATTGGAAGAGAATGAACT
 CATGCTACTGTTCTTGACAAATGCTAGCATGGGAAAAGAAGCTTATGACGAAGTCAAG
 GCTGGCGAATCATGAAATCGAGTACCAGAAAAAGTTGCTCAITTAATCGGGTGAAG
 AAACGAGGTGGCCACTTCGATTCTTAGAGAGAGCTAAAGCAGCAGTAAGTCATTTGCAT
 ACAAGATATATAGTTGATATGCAATCCATGGACTCCACAGTTTCAGAAATCAATCGTCTT
 AGGGATGAACAATACTAACCTAAAGCTCGTTCACCTTGTGAGGCGATGGGGAAGATGTGG
 GAAATGATGCAATACATCATCAAGACAAGCTGAGATCTCAAAGGTGTTGAGATCTCTA
 GATGTTTTCACAGCGGTGAAAGAAACAAATGATCATCATCAGAACGCACCATCCAGCTC
 TTGGCAGTGGTTCAAGAATGGCACACGCAGTTTTCAGGATGATAGATCATCAGAAAGAA
 TACATAAAGCACTTGGCGGATGGCTAAAGCTAAATCTCATCCCTATCGAAAGCACACTC
 AAGGAGAAAGTATCTTCGCTTCTCGAGTTCCCAATCCCGCAATCCAAAACTCCTCCAC
 GCTTGGTATGACCGTTTAGACAAATCCCCGACGAAATGGCTAAAAGTGCCATAATCAAT
 TTCGAGCGGTGTAAGCACGATAATGCAGCAGCAAGAAGACGAGATAAGTCTCAGAAAC
 AAATGCCAAGAGACAAGAAAAGAAATGGGAAGAAAAATTAGACAGTTTGAGGATTGGTAC
 CACAAATACATCCAGAAGAGAGGACCGGAGGGGATGAATCCGATGAAGCGGATAACGAT
 CATAATGATGAGGTGCTGTGAGGCAATTCATGTAGAACAATAAAGAAGAGGTTGGAA
 GAAGAAGAAGAGCTTACCATAGACAAAGCCATCAAGTTAGAGAGAAGTCACTGGCTAGT
 CTTGCACTCGCTTCCCGAGCTTTTTCAGGCAATGTCCGAGGTTGCGTATTCTGTTTCG
 GATATGTATAGAGCTATAACGTATGCGAGTAAGCGGCAAGCCAAAGCGAACGGCATCAG
 AAACCTAGCCAGGGACAGAGTTCGTAAGAACTAATGTAAGATCAGAGTAATGTCTTCTTC
 TTCTTTGATCTTGAATATTTAAGCACACATACATACAACGTATAGCTAAATCTTTATC
 ATTGCTTTCTTATATTAAGGTTTTGGCTTTTGTAGAAGGTTTCTTACATATGAGATTCA
 TATAGTGTGATTCTTAAGGAACGTCTGTTGAGTAATAAGAAAGTTGTGTATTGAAA
 TAGAGTTGCATTTGTTAATTTG

>G1089 Amino Acid Sequence (domain in AA coordinates 425-500)

MGCAQSKIENEEAVTRCKERKQLMKDAVTARNFAAAHSAYAMALKNTGAALSDYSHGEF
 LVSNNHSSSSAAAAIASTSSLPTAISPLPSSTAPVSNSTASSSSAAVPQPIPDTLPPPPP
 PPPLPLQRAATMPENGRSGGGHAGSGLNGIEEDGALDNDDDDDDDDDSEMNDRDLIR
 KRSRSGGSTRGNRTTIEDHHLQEEKAPPPPLANSRPIPPRQHQHQHQHQQQQPFYDYF
 FPNVENMPGTTLEDTPPQPQPQPTRPVPPQPHSPVVTEDDEDEEEEEEEEEETVIER
 KPLVEERPKRVEEVTIELEKVTNLRGMKSKGIGIPGERRGMMPVTATHLANVFIELDD
 NFLKASESAHDVSKMLEATRLHYHSNFDNRGHIDHSARVMRVITWNRSFRGIPNADDGK
 DDVDLEENETHATVLDKLLAWKKLYDEVKAGELMKIEYQKKVAHLNRVKKRGHSDSLE
 RAKAAVSHLHTRYIVDMQSMDSVSEINRLRDEQLYLKLVHLVEAMGKMWEMMQIHQRQ
 AEISKVLRSLDVSQAVKETNDHHHERTIQLLAVVQEWHTQFCRMIDHQKEYIKALGGWLK
 LNLPIESTLKEKVSSPPRPVNPAPIQKLLHAWYDRLDKIPDEMAKSAIINFAAVVSTIMQ
 QQEDEISLRNKCEETRKLGRKIRQFEDWYHKYIQKRGPEGMNPDEADNDHNDEVAVRQF
 NVEQIKKRLEEEEEEYHRQSHQVREKSLASLRLPELFQAMSEVAYSCSDMYRAITYAS
 KRQSQSERHQKPSQGQSS*

>G1093 (1..531)

ATGGGTTATCCGGTGGGGTACACTGAGCTCCTCCTCCCAAGAATCTTCCTTCACTTACTC
TCTCTCTTAGGCTTAATACGAACACTCATAGACACGGGTTTTCGGATATTGGGTCTACCC
GACTTTCTCGAATCCGACCCGGTTTCATCGTCATCGTCATGGCTGGAACCACCGTATATG
TCCACGGCGGCGCATCATCACCAAGAAAGCTCATTTTCTTCCAGTGGCGGCGAGGCTA
GCTGGAGAAATCTTGCCCGTCATCAGATTCTCGGAGCTAACTCGACCCGGATTTCGGATCC
GGATCCGATTGCTGCGCGGTGTGCTCCACGAGTTCGAGAACGATGACGAGATCCGACGG
CTGACGAATTGTCAACACATATTTACCGGAGCTGTTTAGACCGTTGGATGATGGGTAT
AATCAGATGACGTGTCCACTTTGTAGAACGCCGTTTATTTCTGATGAGTTACAAGTTGCT
TTTAACCAACGAGTTTGGTCTGAATCTGAACTTCTCGCAGAATCAAATTAG

>G1093 Amino Acid Sequence (domain in AA coordinates: 105-148)

MGYPVGYTELLLPRIFLHLLSLGLIRTLIDTGFRLGLPDFLES DPVSSSSSWLEPPYM
STAAHHHQESSFFPVAARLAGEILPVIRFSELTRPGFGSGSDCCAVCLHEFENDDEIRR
LTNCQHIFHRSCLDRWMMGYNQMTCPLCRTPPFISDELQVAFNRVWSESELLAESN*

>G1127 (191..1351)

GACAGACTCTCTGTATGTGTGCGAGAAGCGAGAAGCGAGAGAGAGAGAGAGAGATTG
TTAGCTCACACGCTTTCTCTATTTTCTCGGAATTCACAAAACAGAAAGTTTCATCCTTTA
CGAGAAATTAAGCCGAAAGAAACAATCTTTGAGTTTGATTTCTTCTTCTTCTCTCTCT
CTCTGCTCTAATGGATTCCAGAGACATCCACCGTCACATAACCAGCTTCAACCACCACC
GGGAATGTTAATGTCTCATTACCGTAACCTTAACGCCGCGCTTACCATTAAATGGTTCC
CACTTCCACATCTCAACCGATTCAACACCCTCGTCTTCTTTTGGCAATCAACAACAATC
TCAAACGTTTCATCAGCAGCAACAACAACAATGGATCAGAAGACTCTTGAATCTCTTGG
ATTTGGTGATGGATCACCTTCTTCTCAACCGATGCGATTTCGGATCGATGATCAGAATCA
GCAACTGCAAGTGAAGAAGAAGCGAGGAAGGCCGAGAAAGTATACCTCTGATGGTAGCAT
TGCTTTAGGTTTAGCTCTACGTCTCTCTCTCTCTGACGTTCTAATTCTTACCGGTGA
GGGTGGTGTGGAGATAGTGGTGGAAATGGAACTCTGTTGATCCACCTGTTAAACGTAA
CAGAGGAAGGCCTCTGGTTCTAGTAAGAAACAGCTTGATGCTTTAGGAGGAACCTCAGG
AGTTGGGTTTACACCTCATGTATTGAAGTGAACACAGGAGAGGACATAGCGTCAAAGGT
GATGGCTTTTTCGGATCAAGGGTCAAGAACAATTTGTATTCTCTCTGCAAGTGGTGCAGT
TTCTAGAGTGATGCTTCGTCAAGCTTCTCATTCTAGTGAATCGTTACTTATGAGGGACG
ATTTGAGATCATTACTCTCTCAGGCTCAGTCTTGAATTATGAGGTAAATGGTTCCACCAA
CAGAAGTGGTAACCTGAGTGTGGCTTTGGCTGGACCTGATGGCGGCATCGTAGGTGGCAG
TGTAAGTTGGTAATCTAGTAGCTGCAACACAAGTCCAGGTGATAGTGGGAAGCTTTGTTGC
AGAAGCAAAGAAACCGAAACAAAGTAGTGTAAACATTGCTCGGGGCGAGAATCCTGAACC
GGCTTCAGCCCGCGCTAACATGTTGAACCTTTGGATCAGTCTCTCAAGGACCATCGAGCGA
GTCATCAGAAGAGAATGAGAGCGGTTCTCTGCAATGCACCGTGACAATAATAATGGGAT
ATATGGAGCTCAACAACAACAACAACAACCTCTTCATCCTCATCAGATGCAAAATGTA
CCAACATCTTTGGTCTAATCATGGTCAATAAAATGAAGCGGAATTAATTTGTTCCGTT
TTGGTTACGGTTATGGTTTGATTTCTT

>G1127 Amino Acid Sequence (domain in AA coordinates: 103-110, 155-162)

MDSRDIPPSHNQLQPPPGMLMSHYRNPNAASPLMVPTSTSQPIQHPRLPFGNQQSQT
HQQQQQMDQKILES LGFGDGPSSQPMRFGIDDQNQLQVKKRGRPRKYTPDGSIALG
LAPTSPLLSAASNSYGE GGVGDSGGNGNSVDPPVKRNRGRPPGSSKKQLDALGGTSGVGF
TPHVIEVNTGEDIASKVMAFS DQGSRTICILSASGAVSRVMLRQASHSSGIVTYEGRFEI
ITLSGSLVNLNYEVNGSTNRSNL SVALAGPDGGIVGGSVVGNLVAATQVQVIVGSFVAEAK
KPKQSSVNIARQNPESAPAPANMLNFGSVSQGPSSSESSEENESGSPAMHRDNNNGIYGA
QQQQQQQLPHPHQMOMYQHLWSNHGQ*

>G1131 (57..758)

TCGACTCCTCTCTGATTGCTTCACTTCTTCTTTACTACAGGTTTCAGCTCCTCAATGT
CCATGGATTGCTTAAGCTACTTCTTTAACTACGATCCTCCTGTCCAGCTCCAGGATTGCT
TTATTCCCAGATGGATATGATTATCCCTGAAACCGATAGTTTCTTCTTCCAATCTCAAC
CGCAACTGGAGTTTCATCAGCCATTGTTTCAAGAAGAAGTCCTTCACAGACCCACTTTC
ACCTTTCTGCCACAGTTTCTTCTCCGCAAGAAATCTTCTCCCTAACCTTAAACG
AAATCTTCAACGAAACACACGACCTCGATTCTTCTCTCCCAACGCAAAACGCCAGAGAC
TTGTTAACTCCAGCTACAATTGTAACACTCAAAACATTTCAGAGCCGTAACCCGAATT
TCTTCGACCTTTTCGGCGACACTGATTTCTGTTCCAGAATCTTGTACCTTCCAGGAGTTTC
GAGTTCCGATTTCTCTTTAGCTTTCAAGGTAGGCCGGGAGATCAAGATGACTCAAAGA

AACCGACGCTTTTCATCTCAGAGCATCGCGGCTAGAGGGAGGAGAAGAAGAATTGCAGAGA
AGACTCACGAGCTCGGAAAACCTCATCCCCGGTGGCAATAAACTTAACACCGCCGAGATGT
TCCAAGCCCGCGCTAAGTATGTCAAGTTTGTGAGAGTCAAGTTGGGATTCTCCAACCTGA
TGCAGACCACAAAGAAGGTAATAACCAACCCCAATAAGAACTTTATCATCCAATTGAAA
CTCTAATCGTGTCTTCTCACAAGCTTCTTAATTTGTTTACGCAGGGTAGCTCTAATGTGC
AAATGGAACTCAGTATTTGCTTGAATCGCAAGCAATCCAGGAGAAGTTATCAACAGAGG
AAGTGTGTTTGGTACCGTGTGAAATGGTTCAAGATCTAACAACTGAAGAAACCATTTGCA
GAACCCCGAATATTTCTCGAGAAATCAACAAGTTACTGTCTAAACATCTGGCTAACTAGT
TTTAGTTTCAAGCCTGAAGTTCTCTATGCCATAAATTTGTGTCTGTTATCGTTGTTTGTGTC
TTCTTAGTTAGTGTCTTGTCTGTTGATTTAGGGGCTAATTATCCTGGTTAATCTCCTCT
TAACTGGGAA

>G1131 Amino Acid Sequence (domain in AA coordinates: 173-220)
MSMDCLSYFFNYDPPVQLQDCFIPEMDMIIPETDSFFFQSQPQLEFHQPLFQEEAPSQTH
FDPFCDQFLSPQEIFLPNPKNEIFNETHDLDFLPTPKRQRLVNSSYNCTQNHQSRNP
NFFDFPFGDTDFVPESCTFCGAEVDFSLAFKVGREGDQDDSKKPTLSSQSIARGRRRRIA
EKTHELGKLIIPGKNLNTAEMFQAAKYVKFLQSQVGILQLMQTTKKVITNPK*

>G1145 (243..1142)
GTGATTTCTCTCTGCCATTTCTTCGATTTGATTTCTGGGTTCTCTTCTCTCGTCTCTC
TTCTGCATGTTTCGCCACTCTACCTTAGAAAAAGGTTACTTTTCGCCTCCGATTTAGGCT
CGATTTGATGAATTCGTCGTCGTGGCTATTTATCAAATTGAGCATTAGGGTTTCTGAT
TTGTGGGTTTCAAAATGTTTTATCTATCTGTCTTGTGTTTTTGTCCGCTACAAAAGC
CTATGGATTCTCAGAGGGGTATTGTTGAACAAGCTAAATCTCAGTCTTGAATAGGCAAA
GCTCTCTTTACAGCTTAAACACTTGATGAGGTTCAAATCACTTGGGGAGTTCTGGTAAAG
CTCTGGGAAGCATGAACCTTGATGAGCTTTTGAAGAGTGTCTGTTCTGTTGAAGCTAATC
AGCCATCGTCTATGGCTGTCAATGGTGGAGCAGCTGCTCAGGAGGGTCTTTCTCGCCAGG
GGAGTTTGACTTTGCCTCGGGATCTCAGCAAAAAGACTGTTGATGAGGTTTGGAAAGACA
TTCAGCAGAATAAGAATGGAGGTAGTGCTCATGAGAGGAGGATAAGCAGCCTACACTTG
GGGAAATGACGCTTGAAGACCTGTTGTTGAAGCAGGAGTGGTCACTGAGACTATCCCTG
GTTTCAACCATGATGGTCTGTTGGTGGTGGTAGTGCTGGTTTCAAGTGTGTTTAGGGC
AAAACATTACTCAAGTTGGCCCATGGATTCAATATCATCAGCTCCCATCAATGCCACAGC
CTCAAGCATTATGCCCCTATCCGTTTTCAGATATGCAAGCAATGGTGTCTCAGTCTTCTT
TGATGGGTGGTTTGTGATACACAACTCCTGGAAGGAAGAGGGTAGCTTCAGGAGAAG
TTGTAGAGAAGACTGTAGAGAGGAGGCAGAAGAGAATGATAAAGAACAGAGAGTCTGCTG
CTCGTTCGCCGAGCTAGGAAACAGGCTTACACTCATGAGCTAGAGATCAAAGTTTACCGGT
TAGAAGAAGAAAACGAAAGACTCAGGAAGCAAAAGAGGTGGAAAAATCCTCCCAAGTGT
ACCACCGCTGATCCCCAAGCGGAGCTCCGACGGACAAGCTCGGCTCCTTTCTGATCTCT
AAACTCTTTTGTCTTTTCTTTTCTTCTTCTGTGTCGGTTCACTTATAAAAAAGAGA
GGAAAACAGCTTTGTTTCTTTGTACATTCCGTAGACTTTCTTGACTTGGAGCAATTCTGT
TAACTTTAAATATTCTCGAGTTATTGTAGTAGCAGACTAGCAGCAGTAATGGTTTTCAT
GAGTCCGATTGAAATTCAGAGATTGAACAGGAAAAAA

>G1145 Amino Acid Sequence (conserved domain in AA coordinates: 227-270)
MDSQRGIVEQAKSQSLNRQSSLYSLTLDEVQNLGSSGKALGSMNLDELLKSVCSVEANQ
PSSMAVNGGAAAQEGLSRQGSLLTLPRLDSKKTVDVWKDIQQNKNGGSAHERRDKQPTLG
EMTLEDLLLKAGVVTETIPGSNHDGPVGGGSAAGSAGLGQNTQVGPWQIYHQLPSMPQP
QAFMPYPVSDMQAMVSQSSLMGGLSDTQTPGRKRVASGEVVEKTVERRQKRMKNRESAA
RSRARKQAYTHELEIKVSRLEENERLRKQKEVEKSSQVYHRLIPSGSSDGQARLLSDL*

>G1229 (123..1217)
TTTGGGCGGGTCTTCTTTCCCTAAATCTTTCTTTTATTTTGTGTTTAAAAAAAATC
CAACCATAAGACAAAACAACGAACGAGGAAGAGAGAGAGAAGGATATATCTCTAATCA
CGATGCAGGAGATAATACCGGATTTCTTGAAGAGTGTGAATTTGTGACACTTCACTAG
CCGGAGATGATCTATTGCCATCTTAGAGAGTCTTGAAGGTGCCGAGAGATATCTCCGA
CAGCTGCATCTACACCTAAAGATGGAACCAAGTTCCAAGGAGTTAGTTAAGGATCAAG
ATTATGAAAACCTCATCTCCTAAGAGGAAAAAGCAAAGACTAGAAACCAGGAAAGAAGAGG
ACGAAGAAGAAGAAGACGGAGACGGAGAAGCAGAAGAAGATAATAAGCAAGATGGGCAAC
AAAAGATGTCTCATGTAACCGTGGAACGTAACCGGAGAAAGCAAATGAACGAGCACTTAA
CCGTTTTCGTTCTCTTATGCCCTTGTCTACGTCAAACGGGGGACCAAGCATCGATCA
TAGGAGGAGTTGTGGAGTACATAAGCGAGTTACAACAAGTTCTCCAATCTTTGGAAGCCA

AGAAACAACGTAAACCTACGCCGAAGTCCTAAGCCCGAGAGTTGTCCCGAGCCCTCGTC
 CTTACCCGCCTGTTCTAAGCCCAAGAAAACCGCCTCTTAGCCCGCGCATCAACCACCACC
 AGATTCACCACCACCTACTTCTCCCTCCCATAAGTCCTCGAACACCTCAGCCAACAAGCC
 CATACCGGGCCATTCCACCGCAACTACCACTCATCCACAGCCTCCGCTTCGCTCTTACA
 GCTCATTTGGCCAGTTGTCAGAGCTTAGGAGATCCACCTCCATACTCTCCTGCTTCATCTT
 CTTTCATCTCCTTCAGTTAGTAGTAACCATGAGAGTAGTGTGATCAATGAGCTTGTGCTA
 ACTCAAAATCGGCTTTTGGCTGATGTGGAAGTGAAGTTTTCAGGAGCTAACGTGCTGCTCA
 AAACGGTGTGCGATAAGATCCCGGGACAAGTTATGAAGATAATTGCTGCTCTTGAAGATT
 TGGCTCTTGAGATTCTTCAGGTTAATATTAACACCGTCGACGAAACCATGCTTAATTCTT
 TCACCATCAAGATTGGAATTGAGTGCCAACTAAGTGCGAAGAAGTGGCTCAACAAATTC
 AGCAACATTCTGCTAGTAAAGAAGGATTTAATATAGCTTCGTATAAACCTTAACGAGAG
 AGCAGTAGCTACCTACCTTCTCTCCTTAGTATCCCTTTAATTATCTTTTCAGTTTTCGTC
 AAAGATATGGAGTTTAAAAAATAAAATTTGTTATCTAAAGTTTAAATCAAATATTGATTA
 ATTATACTAATATAGGTATAAGTGAGTTTAAAGATTATCAGCTTCATAACAGCCATCG
 TCATGTTTACTTTCTTTTAAATTTTAGAATTTAGACGTACTCCTACCATGTAATTTTATT
 TCTGTCTATACATCAAGCATTGTAGCTGTAATTGCATATGAATGAACAATAGTGTATGAG
 TGATCTCATGAATAATATTCTTCTTGCAACACAAAAA

>G1229 Amino Acid Sequence (domain in AA coordinates: 102-160)
 MQELIPDFLEECEFDVSLAGDDLFAILESLEGAGEISPTAASTPKDGTSSKELVKDQD
 YENSSPKRKKQRLERKEDEDEEEDGDGEAEDNKQDQGQKMSHVTVERNRRKQMNHLT
 VLRSLMPCFYVKRGDQASIIIGVVVEYISELQQVLQSLQLEAKKQKTYAEVLSRVVPSRP
 SPPVLSPRKPLSPRINHQQIHHHLLLPISPRTPOPTSPYRAIPPQLPLIPQPLRSYS
 SLASCSSLGDDPPYPSPASSSSSPSVSSNHESVINELVANSKSALADVEVKFSGANVLLK
 TVSHKIPGQVMKIIAALEDLLEILQVNINTVDETMLNSFTIKIGIECQLSAEELAQQIQ
 QTFC*

>G1246 (1..1746)

ATGATCATGTACGGAGGAGGAGGAGCAGGGAAGGACGGTGGATCCACCAATCACTTATCA
 GACGGAGGAGTGATATTGAAGAAAGGTCCATGGACGGCGGCGGAAGATGAGATACTTGCT
 GCGTACGTTAGAGAGAACGGTGAAGGGAATTGGAACGCCGTTTCAGAAAAACACAGGTTG
 GCTCGTTGCGGCAAAAGCTGCCGTCTTCGATGGGCCAATCACCTCCGACCAAATCTGAAA
 AAAGGCTCTTTACCGGTGACGAAGAAGCTCTCATCATTACAGCTTCATGCTCAGCTTGGT
 AACAAATGGGCTCGCATGGCTGCTCAGTTACCGGAAGAACAGACAACGAGATTAGAAC
 TATTGGAACACGAGATTGAACGACTTCTTCGCCAAGGACTTCCTCTTTATCCTCCAGAT
 ATTATCCCTAACCATCAACTCCATCCACATCCACATCATCAACAACAACAGCAACATAAC
 CATCATCATCATCATCAACAACAACAACATCAACAAATGTATTTTCAACCACAA
 TCTTCACAACGAAACACACCATCATCTTCCCCTCTTCCATCTCCAACACCAGCAAACGCA
 AAGTCCTCATCATCTTCACTTTTCATACACGACTGTAACCTCCTCCATCCACTTAGC
 CCTCACACTCCAAACACACCATCTCAACTCTCTTCCACACCGCCTCCACCACCACTTCC
 TCTCCTTTATGTTCCCTTCGCAACAACCAATACCCGACCTTCCCCTCTTGGCCCTCCCG
 CGTTCCCAAAATCAACAACAACAACAACGGAATTTCACTTTCCCTAGACCTCCACCTCTC
 CTTCAACCGCCTTCATCACTCTTCGCAAAACGTTACAACAATGCTAACACTCCTCTTAAT
 TGCATCAACCGCGTCTCAACCGCACCATTTTCCCCTGTTTCAAGAGACTCCTACACTTCC
 TTTCTTACATTGCCTTACCCTTCCCCAACCGCTCAAACCGCTACTTACCACAATACTAAT
 AACCCCTTACTCTTCTCTCCTTCTCTTTAAACCCCTCTTCTTCTTCTTACCCCTACA
 TCAACTTCTTCCCCAAGCTTTCTTCACTCCCATTAACCTCTTCCACCTCATTTCAT
 ACCAACCCAGTTTACTCCATGAAACAAGAGCAGCTCCCTTCAAACCAAATCCCCAAATA
 GATGGCTTCAATAACGTCAACAACTTTCACAGACAACGAGAGACAGAATCATAACCTTAAC
 AGTTCCGGTGCTCAFGAAGAAGTAGTAGCTGCAGCCTCTTAGAGGATGTCTTGAAGAG
 GCCGAAGCTTTAGCCTCTGGAGGCAGAGGCCGACCTCCAAAACGAAGACAACCTCACAGCT
 TCTCTTCCGAACCACAACAACAACCAACAACAACGACAACCTTCTTCTCGGTTAGTTTC
 GGACATTATGATCTTCTGACAACTTATGTTCTTGAAGATTTGAAATCAAAGGAAGAA
 GAGTCTCTTCAAATGAACACAATGCAGGAGGACATAGCTAAGCTTCTTGATTGGGGAAGT
 GATAGTGGAGAGATCTCTAATGGACAATCATCTGTTGTCACTGACGACAATCTTGTCTT
 GATGTTTCATCAATGACTTCACTTATCCCGGCTGATTCTACAGCCGTCGTAGCCGCAACA
 AACGACCAACAACAAGAATAATAACAATAATTGTTCTGGGATGACATGCAGGGAATA
 AGGTAG

>G1246 Amino Acid Sequence (domain in AA coordinates: 27-139)

MIMYGGGGAGKDDGGSTNHLSDGGVILKKGPWTAAEDEILAAVYRENGEGNWNNAVQKNTGL
ARCGKSCRLRWANHLRPNLKKGSFTGDEERLI IQLHAQLGNKWARMALPGRSTDNEIKN
YWNTRLKRLRLRQGLPLYPPDI IPNHLHPHPPHHQQQQQHNHHHHHHQQQQQHQQMYFQPQ
SSQRNTPSSSPLPSPTPANAKSSSSFTFHTTTANLLHPLSPHTPNTPSQLSSTPPPPPLS
SPLCSPRNNQYPTLPLFALPRSQINNNGNFTFPRPPPLLPSSSLFAKRYNNANTPLN
CINRVSTAPFSPVSRDYSFTLTPYPSPTAQATYHNTNPNYSSSPSFLNPSSSSYPT
STSSPSFLHSHYTPSSSTFHTNPVYSMKQEQQLPSNQIPQIDGFNNVNNFTDNERQNHNLN
SSGAHRRSSSSCLLEDVFEEAEALASGGGRPPKRRQLTASLPNHNNTNNDNFFSVSF
GHYDSSDNLCSLQDLKSKEESLQMNMQEDIAKLLDWGSDSGEISNGQSSVVTDDNLVL
DVHQLASLFPADSTAVVAATNDQHNKNNNNCSWDDMQGIR*

>G1255 (138..1388)

CAGCTCAAACCTCTCTAGGACTACACTAAATCTAACTTTTTCAGAGAGCAAAGATTCAA
TAATTGAGATTGATCTCAAACCAAAGCTCTCGTGCTCTTGCTGTTGATGTTGGTTGTGT
AGACTTTGTATACAATGATGAAAAGTTTGGCGAATGCTGTTGGAGCGAAGACGGCGAGGG
CTTGCGACAGCTGCGTGAAGAGACGTGCACGGTGGTACTGCGCGGCCGACGATGCTTTTC
TTTGCCAGTCTTGCGACAGTTTGGTCCATTAGCAAACCTCTTGCTCGCCGCCACGAGA
GAGTCCGTTTGAAGACGGCTAGCCCGCGGTGCTAAAGCATAGCAACCACTCATCAGCTT
CTCCTCCACATGAGGTCGCCACGTGGCATCACGGGTTTACTCGTAAAGCTCGAACGCCAC
GTGGCTCTGGTAAGAAAAACAATTCTGCGATATTTTCATGACTTGGTTCTGATATTAGTA
TTGAGGATCAGACAGACAACTATGAGCTTGAAGAGCAGCTGATCTGTCAAGTGCCGGTTC
TAGATCCGTTGGTGTCTGAGCAGTTCTTGAACGATGCTGTTGAGCCCAAGATCGAGTTTC
CTATGATCAGAAGTGGTTGATGATCGAGGAGGAGGAAGACAACGCTGAAAGTTGTCTTA
ATGGATTTTCCCGACCGACATGGAGCTTGAGGAGTTTGTGCTGCTGAGCTGGAGACTCTGC
TCGGTTCGCGGGTTAGACACGGAGTCGTATGCCATGGAGGAGCTAGGTTATCTAATTCAG
AGATGTTCAAAATCGAAAAAGATGAGATTGAAGAAGAAGTAGAAGATAAAAGCCATGA
GCATGGATATATTTGATGATGATCGAAAAGACGTGGATGGAACAGTACCGTTTGAGCTAA
GCTTTGATTACGAGTCGTACACAAGACGTCCGAAGAAGAGGTAATGAAGAAGCTTGAAA
GTAGTGTGAATGTGTGTGAAGTGAAAGAGGAAGAACATAAGAATGTTCTGATGCTAA
GATTAACTATGACTCGGTGATATCCACTTGGGGAGGTCAAGGTCCACCGTGGAGTTCAG
GAGAGCCACCGGAACGAGACATGGACATCAGCGGTTGGCCAGCTTTTCCATGGTGGAGA
ATGGAGGAGAAAGTACTCATCAGAAGCAATACGTTGGTGGATGTTTACCATCAAGTGGGT
TTGGAGATGGAGGTAGAGAAGCTAGAGTTTCGAGATACAGAGAGAAGAGGAGGACAAGGT
TGTTTTCTAAGAAGATACGGTACGAGGTACGTAAATTGAATGCAGAGAAAAGACCAGAA
TGAAAGGAAGATTCTGTAAGAGAGCCTCGCTCGCTGCTGCTTACCATTAGGTGTTA
ATTACTGAATAGTTAATATCTATTCTATGTTATATCTCACTTTACAAATTTCCGTGAATCT
TTTTTCTTCTGAAACAACAGAGTTATTTTGGCACTTAATTGTGCTTTGAGGACTTGAT
GTACATAGAAGTAACCAATAATAATGTGACTTTTACTA

>G1255 Amino Acid Sequence (domain in aa coordinates: 18-56)

MKSLANAVGAKTARACDSCVKRRARWYCAADDAFLCQSCDSLVSANPLARRHERVRLKT
ASPAVVKHSNHSSASPPHEVATWHHGFTRKARTPRGSGKKNNSSIFHDLVPDISIEDQTD
NYELEEQLICQVPVLDPLVSEQFLNDVVEPKIEFPMIRSGLMIEEEDNAESCLNGFFPT
DMELEEFADVETLLGRGLDTESYAMEELGLSNSEMFKIEKDEIEEEVEEIKAMSMDIFD
DDRKDVDGTVPFELSFDYESHKTSEEEVMKNVSSGECVVKVEEEHKNVLMRLNLYDS
VISTWGGQGPWSSGEPERDMDISGWPAFMSVENGGESTHQKQYVGGCLPSSSGFGDGG
EARVSRVREKRTRLFSKKIRYEVRLNAEKRPKGRFVKRASLAAAASPLGVNY*

>G1304 (1..978)

ATGGGGCGATCACCATGTTGCGATGAGAATGGTCTAAAGAAAGGCCATGGACACAAGAG
GAGGATGATAAACTGATAGATCACATTCAAAAACATGGCCATGGCAGCTGGAGAGCTCTT
CCAAAGCAAGCCGGTTTAAACCGATGCGGAAAGAGTTGTAGATTAAGATGGACCAACTAC
TTGAGACCTGACATCAAGAGAGGAAATTTCACTGAAGAGGAAGAACAACCTATTATCAAC
CTCCATTCCCTTCTTGGAACAAGTGGTTCGTCGATAGCCGGTAATCTTCTGGAAGAAGC
GACAATGAAATAAAAACTATTGGAACACACATTTGAGAAAGAACTTCTCCAAATGGGG
ATTGATCCGGTGACCCATAGGCCAAGAACCGACCATCTAAACGTTTTAGCAGCTCTCCCG
CAGCTTATAGCCGCCGCAAAATTTCAACAGCCTCTTGAATCTCAACCAAAATGTGCAACTG
GATGCAACAACCTCTTGCTAAAGCTCAACTGCTACACACTATGATTCAAGTCTTAGCACC
AATAACAACACCACCAATCTTCTTTTCTTCTCATCAACTATGCAAAACAGTAACACCAAT
CTCTTTGGCCAAGCTTCTTACTTAGAGAACCAAAATCTTTTGGTCAGTCTCAAACTTC

TCTCACATTCTTGAGGATGAGAATTTGATGGTCAAAACCCAAATTATTGATAACCCCTTTG
GACTCTTTTCTTCCCCATACAACCCGGTTTTCAAGATGATCATAATTCACCTCCCTCTA
TTGGTTCCGGCGTCTCCTGAAGAATCTAAAGAACTCAAAGGATGATCAAGAACAAGAC
ATCGTCGATTACCATCATCATGATGCTTCAAACCTTCATCATCAAACCTCAACGTTTACA
CAAGATCATCATCACCATTGGTGTGACACTATTGATGATGGAGCAAGTGATTCTTTTGG
AAAGAGATAATAGAGTAA

>G1304 Amino Acid Sequence (conserved domain in AA coordinates:13-118)

MGRSPCCDENGLKKGPTQEDDKLIDHIQKHGHSWRALPKQAGLNRCGKSCRLRWNTY
LRPDIKRGNTFEEEEQTIINLHSLGKNSIAGNLPGRTDNEIKNYWNTHLRKKLLQMG
IDPVTHRPRDHLNVLAALPQLIAANFNLSLLNLNQNVLQDATTAKAQLLHTMIQVLST
NNNTTNPSFSSSTMQNSNTNLFQASYLENQNLFQSQNFSHILEDENLMVKTOI IDNPL
DSFSSPIQPGFQDDHNSLPLVLPASPEESKETQRMINKNDIVDYHHHDASNPSSSNSTFT
QDHHHPWCDTIDDGASDSFWKEIIE*

>G1318 (7..849)

AAAAATATGAGGAAGCCAGAGGTAGCCATTGCAGCTAGTACTCACCAAGTAAAGAAGATG
AAGAAGGGACTTTGGTCTCCTGAGGAAGACTCAAAGCTGATGCAATACATGTTAAGCAAT
GGACAAGGATGTTGGAGTGATGTTGCGAAAAACCGAGGACTTCAAAGATGTGGCAAAAGC
TGCCGTCTTCGTTGGATCAACTATCTTCGTCCTGACCTCAAGCGTGGCGCTTCTCTCCT
CAAGAAGAGGATCTCATCATTCGCTTTCATTCATCCTCGGCAACAGGTGGTCTCAGATT
GCAGCAGGATTGCCCTGGTCCGACCGATAACGAGATCAAGAATTTCTGGAACCAACAATA
AAGAAAAGGCTAAAGAAGATGTCCGATACCTCCAACCTTAATCAACAACCTCATCTCATCA
CCCAACACAGCAAGCGATTCTCTTCTAATTCGCGATCTTCTTTGGATATTAAAGACATT
ATAGGAAGCTTCATGTCTTACAAGAACAAGGCTTCGTCAACCCCTTCCTTGACCCACATA
CAAACCAACAATCCATTTCCAACGGGAAACATGATCAGCCACCCGTGCAATGACGATTTT
ACCCCTTATGTAGATGGTATCTATGGAGTAAACGAGGGGTACAAGGGGAACCTTACTTC
CCACCTTTGGAATGTGAAGAAGGTGATTGGTACAATGCAAAATATAACAACCACTTAGAC
GAGTTGAACACTAATGGATCCGGAACGACCTGAGGGTATGAGACCAGTGGAAGAATTT
TGGGACCTTGACCAAGTTGATGAACACTGAGGTTCTTCTGTTTACTTCAACTTCAAACAA
AGCATATGAATATTTTACGTCATCTTATTCTTTTTTCTATTGCGGTTTATACTCAAGAT
TCTTAGCCACACACATAAATGCAAATATATATACATTGTTAGAGAGTATTTTGTATTT
CGTATAATCTTTTCTGACTAGGGCTTGAGCCTTGAGGTCCCATGTACGATTAGTCAATG
TAAACATATATCTTATAATAAATAAATAAAGAAATAATAAGCACATAAAAAAAAAAAAA
A

>G1318 Amino Acid Sequence (domain in AA coordinates: 20-123)

MRKPEVAIAASTHQVKKMKGLWSPEEDSKLMQYMLSNQGQWSDVAKNAGLQRCGKSCR
LRWNYLRPDLKRGAFSPQEEELIIRFHSILGNRWSQIAARLPGRTDNEIKNFWNSTIKK
RLKKMSDTSNLINSSSSPNTASSSSNSASSLDIKDIIGSFMSLQEQGFVNPSLTHIQ
NNPFTGNMISHPCNDDFTPYVDGIYGVNAGVQGEYFPFLECEEGDWYNNINNHDEL
NTNGSGNAPEGMRPVEEFWDLDQLMNTVEVPSFYFNFKQSI*

>G1320 (39..788)

GAAGATCATAAAGATCAAAAGGAGAGAGGTATTAATAATGATGTGTAGTCGAGGCCATT
GGAGACCTGCAGAAGACGAGAAGCTAAGAGAACTCGTCGAGCAATTTGGTCTTCATAATT
GGAACGCCATAGCTCAGAAGCTCTCTGGTCGATCTGGTAAGAGTTGTAGATTGAGATGGT
TTAATCAATTGGATCCTAGGATTAACCGAAACCCCTTCACGGAGGAAGAAGAAGAAAGGC
TTTTAGCGCTCATCGGATCCATGGGAACAGATGGTCTGTGATCGCTAGATTTTTCCTCG
GTGCAACTGATAACGCTGTTAAAAACCATTTGGCACGTCATCATGGCTCGTCTGGCCGAG
AACGGTCCAAGCTCCGTCCACGAGGCCCTTGGCCATGATGGCACGGTGGCTGCGACTGGGA
TGATTGGTAATTATAAAGACTGCGATAAGGAGAGAAGATTGGCAACCACAACCGCTATCA
ATTTTCTTATCAATTCTCTCATATTAATCATTTTCAAGTCTCAAAGAGTCTTTCGACCG
GAAAGATCGGGTTCAGAAATAGTACTACTCAATACAGAAGGAGCAATAGACCAAACTA
AACGACCGATGGAGTTCTACAATTTTCTCCAAGTAAACCGGATTGGAAGATACACGAAT
TGATAGATAATTCAAGAAAAGACGAAGAAGAAGATGTCGATCAAAACAACCGAATTGTA
ACGAGAATTGTGTTCCATTTTTCGACTTTTGTCTGTTGGAACTCTGCCTCTCAGGGTT
TATGTTAATTTGTCGTACCAATGTAAGGTGGACCATATGTTAACTAAAGATAA
TGTAAGAAAGTACTAATCAATTAGAGCTCTGTTTGGCCAAATGTGAAAATTAGTTAAGA
CATCCCAAACATTTTCTGTATTAACACATATAAGGTTGTACTTTTATCAGGTCTAATTTT
CTATTTTATTTTAAAGGATGTTTAAATCAGACCCATAACCATTCGATAAAAAAAAAAAAAA

>G1320 Amino Acid Sequence (domain in AA coordinates: 5-108)
MMCSRGRHWRPAEDEKLRELVEQFGPHNWNIAQKLSGRSGKSCRLRWFNQLDPRINRNPF
TEEEERLLAPHRIHGNRWSVIARFFPGRTDNAVKNHWHVIMARRGRERSKLRPRGLGHD
GTVAATGMIGNYKDCDKERRLATTTAINFPYQFSHINHFOVLKESLTGKIGFRNSTTPIQ
EGAIDQTKRPMEFYNFLQVNTDSKIHLEIDNSRKDEEDVDQNNRIRNENCVPFFDFLSV
GNSASQGLC*

>G1330 (36..959)

GTACCGGCGACCTCTTTGTGGGTCACTCTTCATCAATGGGTGACAAAGGAAGGAGCTTAA
AGATCAACAAGAACATGGAGGAATTCACGAAAGTGAAGAAGAAATGGACGTAAGGAGAG
GTCATGGACAGTTGAGGAAGATTTAGAGCTCATCAATTACATTGCTAGTCATGGTGAAG
GTCGATGGAACCTCTTCGCTCGTTGCGCCGAACCTCAAAGGACCGGAAAAAGCTGCAGAC
TTCGGTGGCTGAACCTATCTCCGACCAGATGTGCGCCGTGGAACATAACCCCTCGAAGAAC
AACTCTTGATTCTTGAACCTTCACACACGTTGGGGCAATAGATGGTCTAAGATTGCACAAT
ATTTACCAGGAAGAACGGATAACGAGATCAAAAATATTGGAGAACACGTGTTCAAAAGC
ATGCAAAACAGCTTAAATGCGACGTGAACAGTCAACAATTTAAAGACACCATGAAGTATC
TTTGGATGCCTCGGCTCGTAGAAAGGATCCAAGCCGCGTCCATCGGGTCTGTTTCCATGT
CATCTTGCGTCACCACCTCCTCAGATCAGTTCGTGATCAACAACAACAACACCAACAACG
TGGATAATTGGCTTTAATGAGTAACCCCTAATGGTTACATCACGCCGGATAATTCCAGCG
TGGCAGTATCTCCTGTATCAGATTGACCGAGTGTCAAGTGAAGTGAAGTGTGAAGA
TTGGTCAGGATGAGAATTTGGTGGATCCAAAATGACATCGCCGAATTATATGGATAATA
GCAGTGGACTATTAAACGGAGATTTTACGAAGATGCAAGATCAAAGTGACCTTAATTGGT
TTGAAATATTAAATGGGATGGTACCAAATTATTCGGACAGTTTTTGAACATTGGAAATG
ATGAAGACTTCTGGCTCTTACAACAACATCAACAAGTCCACGACAATGGAAGCTTCTGAA
TAGACAAGAAGCTATGCGGCC

>G1330 Amino Acid Sequence (domain in AA coordinates: 28-134)
MGDKGRSLKINKNMEEFTKVEEMDVRRGPWTVEEDLELINYIASHGEGRWNSLARCAEL
KRTGKSCRLRWLNLRPDVRRGNITLLEEQLLILELHTRWGNRWSKIAQYLPGRDNEIKN
YWRTRVQKHAKQLKCDVNSQQFKDTMKYLWMPRLVERIQAASIGSVSMSSCVTTSSDQFV
INNNTNNVDNLALMSNPNGYITPDNSSVAVSPVSDLTECQVSSEVWKIGQDENLVDPKM
TSPNYMDNSSGLNGDFTKMQDQDLNWFENINGMVPNYSDFWNIGNDEDFWLLQQHQQ
VHDNGSF*

>G1352 (79..900)

GCGCGATTAAAACTCTCAACTTTTCTCTCAAAATTTCTGATCCTTTGATCCAACAGTTAG
AAGAAGATTTCATCTGATCATGGCCCTCGAAGCGATGAACACTCCAACCTCTTTTACC
AGAATCGAAACGAAAGAAGATTTGATGAACGACGCCGTTTTCATTGAGCCGTGGCTTAA
CGAAACGCTCCAAACGTCAGCGTTCTCACAGCCCTTCTTCGTCTTCTTCTCACC GCCT
CGATCTCGACCCAAATCCAGAATCAAGATCTTACGGAAGAAGAGTATCTCGCTCTTTGT
CTCCTCATGCTCGCTAAAGATCAACCGTCGCAACGCGATTTTCATCAACAGTCGCAATCG
TTAACGCCGCGCCGAGAAATCAAGAACCTTCCGTACAAGTGAACGTCTGTGAAAAAGCG
TTTCTTCTCTATCAGGCTTTAGGCGGTACAAAGCAAGTCACCGAATCAAACCACCAACC
GTAATCTCAACAACCGCCGATGATTCAACAGCTCCGACCATCTCCATCGTCGCCGAGAA
AAACATCCGATTGCTGCCTCCGGAAGATCCACGAGTGTTCATCTGTCATAAAGTGTTT
CCGACGGGTCAAGCTTTAGGCGGTACAAACGTTGTCACTACGAAGGCAACCTCGGCGGC
GGAGGAGGAGGAGGAAGCAAATCAATCAGTCACAGTGAAGCGTGTGAGCACGGTATCG
GAAGAAAGGAGCCACCGTGGATTATCGATCTAAACCTACCGCGCTTACCTGAACCTCAGC
CTTCATCAATCAATCGTCGACGAAGAGATCTTGAGTCCGTTGACCGGTAAAAAACCG
CTTTTGTGACCGATCAGCACCAAGTCATCAAGAAAGAAGATTATCTTTAAAAATCTAA
TACTCGACTATTAAFTCTTGTGTGATTTTTTTCGTTACAACCATAGTTTCATTTTCATT
TTTTAGTTACAAATTTTAAATTGTTCTGATTTGGATTGAAA

>G1352 Amino Acid Sequence (domain in AA coordinates: 108-129,167-188)
MALEAMNPTSSFTRIETKEDLMNDVFIIEPWLKRKRKRQRSHSPSSSSSSPPRSRPKS
QNQDLTEEYLALCLMLAKDQPSQTRFHQQSLSLTPPPESKNLPYKCNVCEKAFPSYQA
LGGHKASHRIKPTVISTTADDSTAPTISIVAGEKHPHIAASGKIHECSIHKVFPTGQAL
GGHKRCHYEGNLGGGGGGGSKSISHSGSVSSTVSEERSHRGFIDLNLPALPELSLHNP
VDEEILSPLTGKPLLLTDHDDQVIKKEDLSLKI*

>G1354 (1..1047)

ATGGAAGTCTCGCACACATTCCTCCCGGTTATCGATTCCATCCGACCGATGAAGAATC

GTTGACTATTATCTCAAGAACAAAGTTGCATTCCCGGGAATGCAAGTTGATGTTATCAAA
 GATGTTGATCTCTACAAAATCGAGCCATGGGACATCCAAGAGTTATGTGGAAGAGGGACA
 GGAGAAGAGAGGGAATGGTATTTCTTTAGCCACAAGGACAAGAAATATCCAAGTGGGACA
 CGAACCAATAGAGCAACGGGCTCCGGATTTTGGAAAGCAACGGGTCGAGACAAGGCCATT
 TACTCAAAGCAAGAGCTTGTGGGATGAGGAAGACTCTTGTCTTTTACAAAGGTAGGGCC
 CCAAATGGTCAGAAATCTGATTGGATAATGCACGAATACCGTCTTGAGACCGATGAAAAT
 GGACCGCCTCATGAGGAAGGATGGGTGGTTTGTGCGCTTTCAAGAAGAAGCTAACCCAG
 ATGAACACACAATCCAAGAACAAATGATGGGATCATCATCAGGCCAAGAATCTAACTGG
 TTCACGCAGCAAAATGGATGTGGGAATGGTAATTACTATCATCTTCTGATCTAGAGAGT
 CCGAGAATGTTTCAAGGCTCATCATCATCACTATCATCATTACATCAGAAATGATCAA
 GACCCCTTATGGTGTCTGACTCAGCACTATTAACGCAACCCCAACTACAATAATGCAACGA
 GATGATGGTTCATGTGATTACCAATGATGATGATCATATGATCATGATGAACACAAGTACT
 GGTGATCATCATCAATCAGGATTACTAGTCAATGATGATCATATAATGATCAAGTAATGGAT
 TGGCAAAACGCTTGACAAGTTTGTGTCTTCTCAGCTAATCATGAGCCAAGAAGAGGAAGAA
 GTTAACAAAGATCCATCAGATAATTCTTCGAATGAAACATTTTCATCATCTCTCTGAAGAG
 CAAGCTGCAACAATGGTTTCGATGAATGCTTCTTCTCTTCTTCCATGTTCTCTCTAC
 TCTTGGGCTCAAAATACACACACGTAA

>G1354 Amino Acid Sequence (domain in AA coordinates: TBD)

MESLAHIPPGRFHPHPTDEELVDYLLKNKVAFPQMVDVIKDVLDYKIEPWDIQELCGRGT
 GEEREWYFFSHKDKKYPTGTRTNRATGSGFWKATGRDKAIYSKQELVGMKRLTVFYKGRA
 PNGQKSDWIMHEYRLTDENGPPHEEGWVVCRAFKKLTMMNYPNPTMMGSSSGQESNW
 FTQQMDVGNNGNYHLPLDLESPRMFQSSSSSLSLHNDQDPYGVVLSTINATPTTIMQR
 DDGHVITNDDDHMMNMTSTGDHHSGLLVNDDHNDQVMDWQTLDFVASFQSLMSQEEEE
 VNKDPSDNSSNETFHLSEEQAATMVSMMNASSSSSPCSFYSWAQNTHT*

>G1360 (1..1257)

ATGGGAGATAGAAACAACGACGGTGATCAGAAAATGGAGGATGTATTGTTGCCCGGATTT
 AGGTTTCATCCAACCGACGAAGAGCTCGTAAGCTTCTACCTGAAGCGGAAGGTTCAACAC
 AACCTCTCTCCATTGAGCTCATAAGACAACCTCGATATCTACAAATATGACCCCTGGGAT
 CTTCCAAAGTTTGCATGACGGGTGAAAAAGAATGGTACTTTTATTGTCCAAGGGACAGG
 AAGTATAGGAACAGCTCGAGGCCAAACCGAGTGACCGGAGCTGGTTTTTGGAAAGCCACG
 GGAACGGACCGGCCGATATACTCGTCAGAAGGAAACAAATGCATAGGTTTAAAGAAGTCC
 TTAGTGTTCTACAAAGGAAGAGCAGCGAAAGGAGTTAAGACTGATTGGATGATGCATGAG
 TTTCTGTTTGCCTTCTCTCTCCGAACCATCTCTCTCTTCTAAGAGATTCTTCGACTCTCT
 GTCTCTCCCAACGATTTCATGGGCTATATGCAGAATCTTCAAAAAGACCAACACAACGACC
 CTAAGAGCTCTCTCTCTCTCTTTTGTCTTCTGTTACCAACAGAAACAAGCACCGACACA
 ATGTCTAACCAAAAGCAATCAACACATACCATTTTTTCTTCAGACAAGATCTCAAACCT
 AGCTCTCACTTCCAGTTTCCCATGAGAAATGAACACTCCCAAACTAGTAATAGTACA
 ACTCCATCCGTTCCTCACTATAAGTCCCTTCTCTTACTTGATTTTCACTTCATACGACAAA
 CCCACCAACGTTTTCAATCCGGTTTCATGTTTAGACCAACAATACCTCACAAATCTCTTT
 CTTGCCACACAAGAAACACAACCTCAGTTTCCAGGCTCCCTCGTCAAAATGAAATCCCA
 TCGTTTCTGCTAAACACGCTTTCAGATTGACCTTCTTGGGAGAATTACAGAGCCATATC
 GACCTCAGCGCAGTGTTGGCCCAAGAGCAATGTCCCCGCTTGTAAAGCCTACCACAGGAG
 TATCAAGAGACGGGATTTCGAAGGAAATGGTATAATGAAGAACATGCGTGTTTCAATGAA
 GATCATCTTGGTGATCATTGCGACACACTTCGGTTTGATGATTTCACTTCAACAATTAAT
 GAGAACCATCGTCATCATCAAGACCTGAAACAGAACATGACATTGCTGGAGAGTTATTAT
 TCTTCTTTATCGTCCATCAATAGCGATTGTCAGCTTGTCTTCTCCAGTACAACCTGA

>G1360 Amino Acid Sequence (conserved domain in AA coordinates: 18-174)

MGRNNDGDQKMEDVLLPGFRFHPHPTDEELVSFYLLKRVQHNPLSIELIRQLDIYKYDPWD
 LPKFAMTGEKEWYFYCPRDRKYRNSSRPNRVTGAGFWKATGDRPIYSSEGNKICGLKKS
 LVFYKGRAAKGVKTDWMMHEFRLPSLSESPSPSKRFFDSPVSPNDSWAICRIFPKTNTTT
 LRALSHSFVSSLPETSTDMSNQKQSNTRYHFSDDKILKPSHFQFHENMNTPKTSNST
 TPSVPTISPFSYLDFTSYDKPTNVFNPVSLDQQYLTNLFLATQETQPPRLPSSNEIP
 SFLNLTSSDSTFLGEFTSHIDLAVLAQEQCPPLVSLPQEQYQETGFEGNGIMKNMRGSNE
 DHLGDHCDTLRFDDFTSTINENHRHQDLKQNMTLLESYSSSLSSINSDLPACFSSTT*

>G1364 (1..537)

ATGGCGGAGTCGCAGGCCAAGAGTCCCGGAGGCTGTGGAAGCCATGAGAGTGGTGGAGAT
 CAAAGTCCAGGTCGTACATGTTCTGTGAGCAAGATAGGTTTCTTCCGATTGCTAACATA

AGCCGTATCATGAAAAGAGGTCTTCTCTGCTAATGGGAAAATCGCTAAAGATGCTAAGGAG
ATTGTGCAGGAATGTGTCTCTGAATTCATCAGTTTCGTCACCAGCGAAGCGAGTGATAAA
TGTCAAAGAGAGAAAAGGAAGACTATTAATGGAGATGATTGCTTTGGGCAATGGCTACT
TTAGGATTTGAAGACTACATGGAACCTCTCAAGGTTTACCTGATGAGATATAGAGAGGGT
GACACAAAGGGATCAGCAAAAAGGTGGGGATCCAAATGCAAAGAAAGATGGGCAATCAAGC
CAAAATGGCCAGTTCTCGCAGCTTGCTCACCAGGTCCTTATGGGAACCTCTCAAGTAAC
TTTCTCTCTCTCTCTTACACTCAAGCAATACGCATCATTCTCTTCTAATTTGTTAA
>G1364 Amino Acid Sequence (conserved domain in AA coordinates: 29-120)
MAESQAKSPGGCGSHESGGDQSPRSLHVREQDRFLPIANISRIMKRGLPANGKIAKDAKE
IVQECVSEFISFVTSEASDKQREKRKTINGDDLWAMATLGFEDYMEPLKVYLMRYREG
DTKGSAGKGD PNAKKDQSSQNGQFSQLAHQGPYGNQVTFPLFSSHSNTHSLIC*
>G1379 (68..622)
CTCTGCCTCTCTCTCTCTCAAAACCCATCTCGAAAGTCTTTCTCTTTTCGAGGGTTTAG
ATCCTCCATGGAAGGCGGCGGAGTTGCTGACGTGGCTGTCCCGGTACGAGGAAGAGAGA
CAGACCTTACAAAGGAATTAGGATGAGGAAGTGGGAAAGTGGGTGGCGGAGATTCGTGA
GCCTAACAGCGCTCTAGGTTATGGCTTGCTCTTACTCTACTCCCGAGCGCGCGCGCG
AGCTTACGACACGGCGGTTTCTATCTTAGAGGACCTACGGCGAGGCTTAACCTTCCCTGA
GCTTCTCTCTGGGAGAAATCTCCGACGAGGATATGTCGGCTGCGACCATCAGGAAGAA
AGCCACGGAGGTCGGTGTCTCAGGTTGATGCTTTGGGCACGGCGGTGCAAAATAACGCCA
CCGTGTTTTTGGTCAGAAATCGAGATAGTGATGTGGATAATAAGAATTTTCATCGGAATTA
TCAAACGGTGAACGAGAAGAAGAAGAAGATGAGGATGACAAGAGATTGAGGAGTGG
CGGCCGTTATTGGATCGGGTTGACTTGAATAAATTACCCGACCCGAAAGCTCCGATGA
AGAATGGGAAAGCAAACATTAAAAATATATAGTTTGGAGCGGTGGCTGTGCTAACGTAC
GCCAACGGCTTGCTTCTACGAATCATTAGCGCCGTTTTATGATTTTTTTTTTTTTTTTT
CATTATCTGAAAATTTAGGGCTTTTTAGTTATTAATTTTTGTTTTGTTTTTCTTTCT
TGCGAGTTTTGCGGTTTTATGGAATTTTAGGCTATTGCTTAACGAAAAAAAAAAAAAAAA
>G1379 Amino Acid Sequence (domain in AA coordinates: 18-85)
MEGGGVADVAVPGTRKRDRPYKGI RMRKWKWVAEIREPNKRSRLWLGSYSTPEAAARAY
DTAVFYLRGPTARLNFPELLPGEKFSDEDM SAATIRKKATEVGAQVDALGTAVQNNRHRV
FGQNRSDVDNKNFHRNYQNGEREEEEDEDDKRLRSGRLLDRVDLNLKLPDPESDBEW
ESKH*
>G1384 (33..977)
GTACATTTTTTTTTTGTAATTTTCAGGAACTCCGATGGCGGATCTCTTCGGTGGTGGCCACG
GCGGCGAGCTTATGGAAGCACTTCAACCTTTTTACAAAAGTGCTTCCACGTCTGCTTCAA
ATCCTGCGTTTGCGTCTCTCAAACGATGCGTTTGCGTCTGCCCCAAACGACCTATTTTCTT
CTTCTCTTACTATAATCCTCATGCATCTTTATTCCCTTCACATTCCACAACCTCTTACC
CGGATATTTATTCTGGATCCATGACCTATCCATCTTCAATTCGGGTGCGATCTTCAACAAC
CCGAAAACCTACCAATCTCAGTTCCATTACCAAAACACTATCACTTACACTCACCAAGACA
ACAACACTTGCATGCTTAACCTTCATTGAGCCGAGCCAAACCGGTTTTATGACCCAAACCGG
GTCCGAGTTTCGGGTTTCGGTTTCAAAACCGGCTAAGCTCTATAGAGGAGTGAGGCAAAGAC
ATTGGGGAAAATGGGTGCGCGGAGATCCGTTTACCCAGGAACCGAACCCGACTTTGGCTCG
GAACATTCGACACGGCTGAAGAAGCCGCTTGGCTTATGATCGCGCCGCTTAAAGCTTC
GTGGTGACTCGGCTCGGCTTAACCTCCAGCTCTCCGATACCAACCGGCTCGTCTCCGT
CTGATACCGGCGAATATGGTCTTATTCAAGCTGCCGTAGACGCTAAACTAGAAAGCCATAT
TAGCTGAGCCGAAGAATCAGCCGGGCAAAACGGAGAGGACGTCGAGGAAACGAGCTAAAG
CCGCGGCTTCTTCACTGAGCAGCCGTCAGCGCCACAACAACATTCCGGGTGCGGTGAAA
GTGATGGGTGCGGGTTCCCGACTTCGGATGTTATGGTGCAGGAGATGTGCCAAGAGCCAG
AGATGCCATGGAATSAATAATTTATGCTCGGCAAGTGTCTTCTTATGAGATAGATTGGG
CTTCAATTTTATCGTGAAAAATTAGGATTCAATTCATTTTTATTCTTTAACTTGTTTG
TATTTCTTTTAACTTTAGGGTTATTAGCTGTGCGTAA
>G1384 Amino Acid Sequence (domain in AA coordinates: TBD)
MADLFGGGHGGELMEALQPFYKSASTSASNP AFASSNDAFASAPNDLFSSSSSYNPHASL
FP SHSTTSYPDIYSGSMTYPSSFGSDLQFPENYQSQPHYQNTITYTHQDNNTCMLNFIEP
SQPGFMTQPGPSSGSVSKPAKLYRGVRRHWGKWVAEIRLPNRNRLWLGTDFDTAEBAAL
AYDRAAFKLRGDSARLNF PALRYQTGSSPSDTGEYGP IQAAVDKLEAILAEPKNQPGKT
ERTSRKRAKAAASSAEQPSAPQOHSGSGESDGSPTS DVMVQEMCQEPMPWNENFMLG
KCPSYBIDWASILS*

>G1399 (261..1475)

AGGTGCAATTTTCTGAAATTAAGATTTCATTCCTCCATGGAAGAAGCTCTGTTTTTATTCT
CTTTAGCTTAGCTTAGCTTCTACTGATCTGTTTTTGGCTACAAAATCCCATCTTTTCTTT
AAAACTCTTTATCTCTGAATCTTGAGTTTCTTGTAAGAAGAAGCAATTTTGAATCTTT
CGTAATCATAAAGATTTCGTGGAGGATCTCTACTGATTTGTGGAATCTCTCACTACAGAA
TCACTTGATCTTATGTCCGGATGGAGGAGAGAGAAGGAACCAACATCAACAACAACATCA
CTAGCAGTTTCGGCTTGAAGCAGCAACATGAAGCTGCTGCTTCTGATGGTGGTTACTCAA
TGGACCCACCACCAAGACCCGAAAACCCCTAACCCGTTTTTAGTCCCACCCACTACTGTCC
CCGCGGCCCGCCACCGTAGCAGCAGCTGTTACTGAGAATGCGGCTACTCCGTTTAGCTTAA
CAATGCCGACGGAGAACACTTCAGCTGAGCAGCTGAAAAGAAGAGAGGTAGGCCGAGAA
AGTATAATCCCGATGGGACTTTGTGCTGACTTTATCGCCGATGCCAATCTCGTCTCTG
TTCCGTTGACGTCGGAGTTTCTTCCAAGGAAACGAGGAAGAGGACGTGGCAAGTCTAATC
GATGGCTCAAGAAGTCTCAAATGTTCCAAATTCGATAGAAGTCTGTTGATACCAATTTGG
CAGGTGTAGGAACTGCTGATTTGTTGGTGCCAACTTTACACCTCATGTACTGATCGTCA
ACGCCGGAGAGGATGTGACGATGAAGATAATGACATTCTCTCAACAAGGATCTCGTGCTA
TCTGCATCTTTTCACTAATGGTCCCATCTCCAATGTTACGCTTCGTCATCTATGACAT
CCGGTGGTACTCTAAGCTTATGAGGCTCGTTTTGAGATTCTCTCTTTGACGGGTTCTGTTA
TGCAAAATGACTCTGGAGGAACCTGAAGTAGAGCTGGTGGTATGAGTGTGCTGCTGAG
GACCAGATGGTGGTCTTTTGGTGGAGGACTCGCTGGTCTCTTTCTGCTGCTGGTCTG
TCCAGGTAATGGTAGGGACTTTTATAGCTGGTCAAGAGCAGTCACAGCTGGAGCTAGCAA
AAGAAAGACGGCTAAGATTTGGGGCTCAACCATCTTCTATCTCTTAAACATATCCGCAG
AAGAACGGAAGCGGAGATTCCGAGAGGCTTAACAAGTCTGTTGCTATTCTGACCAACCA
CTTCATACACGCATGTAACACAACAATGCGGTTTACAGTTACTATACAACTCGGTTA
ACCATGTCAAGGATCCCTTCTCGTCTATCCAGTAGGAGGAGGAGGAGGTGGAGAGGTAG
GAGAAGAAGAGGCTGAAGAAGATGATGAATTAGAAGGTGAAGACGAAGAATTCCGAG
GCGATAGCCAACTCGACAACGAGATTCCGAGCTGATGATGATCATACGGTTTCTTTTCGC
GGATTTGTTAGTTTGTATGGATTTCAGATTTTGGTTGATTGTTTTTATTAACACAGAATG
TTTGAAGCTGCTATCTTTAGGTTCCCATCTCTTGTGATTGTTGAGTATCCTTGTAGA
AACAAACTTACTGTTGCAAACTCTCTTCAAAAAAGTTTCACTTTGCTTTCCCA

>G1399 Amino Acid Sequence (domain in AA coordinates: 86-93)
MEEREGTINNNITSSFGLKQOHEAAASDGGYSMDPPRPNPNPFLVPPTTVPAAATVA
AAVTENAATPFSLTMPENTSAEQLKKRGRPRKYNPDGTLVVTLSMPPISSSVPLTSEF
PPRKRGRGRGKSNRWLKKSQMFQFDRSPVDNLAGVGTADFVGANFTPHVLIVNAGEDVT
MKIMTFSQQGSRAICILSANGPISNVTLRQSMSTSGGTLTYEGRFEILSLTGSFMQNDSSG
TRSRAGGMSVCLAGPDGRVFGGGLAGLFLAAGPVQVMVGTFLAQEQSQLELAKERRLRF
GAQPSISIFNISAEERKARFERLNKSAIPAPTTSTYHVNNTNAVHSYYTNSVNHVKDPF
SSIPVGGGGGEGVEEDDDLEGEDEEFGGDSQSDNEIPS*

>G1415 (60..680)

CCTTATCACTACCAAAAGTCGTCACATAATATCACTTTTCGAGTTATCAACATCCGTACA
TGTATCCATAGAGCCAAAAGTAATGATGGTTGGTGCTAATAAGAAACAACGAACCGTCC
AAGCTAGTTTCGAGGAAAGGTTGTATGAGAGGAAAAGGTGGACCCGATAACGCGTCTTGCA
CTTACAAAGGTGTTAGACAACGCAC'TTGGGGCAAATGGGTGCTGAGATCCGCGAGCCTA
ACCGAGGAGCTCGTCTTTGGCTCGGTACCTTCGACACCTCCCGTGAAGCTGCCTTGGCTT
ATGACTCCGAGCTCGTAAGCTCTATGGGCTGAGGCTCATCTCAACCTCCCTGAGTCCT
TAAGAAGTTACCCTAAAACGGCGTCTCTCCGGCGTCCCAGACTACACCAAGCAGCAACA
CCGGTGGAAAAGCAGCAGCGACTCTGAGTCGCCGTGTTTATCCAACGAGATGTCATCAT
GTGGAAGAGTGACAGAGGAGATATCATGGGAGCATATAAACGTGGATTGCGCGTAATGG
ATGATTCCTCAATATGGGAAGAAGCTACAATGTCGTTAGGATTTCCATGGGTTTCATGAAG
GAGATAATGATATTTCTCGGTTTGATACTTGTATTTCCGGTGGCTATTCTAATTGGGATT
CCTTTTCACTCCCACTTTGAGGTGTCACTAGACTCTCTTAATTGTTAAGTTATCATATA
CAAACTACATATATATACAAATATAGTCACCGTGAAGTATAGTAAATAAACA
CCAGTTACATGTACTTATATATGTGCACATCTATATATGTGGTTTGTCTGTATAGTGTGA
AAGCAGATTCTTACCATATCA

>G1415 Amino Acid Sequence (domain in AA coordinates: TBD)
MSSIEPKVMVMGANKKQRTVQASSRKGMRGKGGPDNASCTYKGVQRQTGKQWVAEIREP
NRGARLWLGTFDTSREAAALAYDSAARKLYGPEAHLNLPESLRSYPKTASSPASQTTSSN
TGGKSSSDSESPCSSNEMSSCGRVTEEISWEHINVDLPVMDSSIWEEATMSLGFPPVHE

GDNDISRFDTCISGGYSNWDSFHSP*
>G1417 (32..1501)

TCTATCTCTATCTATCTCTCTTTGTCTGCAATGGAAGACATATTCAAGATCGCCGTGA
AATTGCGTTCTTACACTCAGGAGAATTTCTCCACGGAGATTCTGACTCAAAGGATCATCA
ACCGAACGAGTCTCCGGTGAACGTCATCACGAGTCGTCTATCAAAGAAGTTGATTTCTT
CGCTGCTAAAAGTCAGCCGTTTGATCTTGGTCATGTGAGAACAACGACGATCGTTGGATC
ATCTGGTTTAAATGATGGATTAGGTTTGGTAAATTCATGTCATGGAACATCAAGCAATGA
TGGCGATGACAAAACAAAACCTCAAATTAGTAGACTGAAGTTGGAGCTAGAGAGGCTTCA
CGAGGAGAATCACAACTGAAGCATTATTAGATGAGGTCAGTGAGAGTTACAACGACCT
CCAAAGAAGAGTTTTTGTAGCAAGACAAACACAAGTGAAGGTCTTCATCATAAACAACA
TGAGGATGTACCTCAAGCTGGTTCCTCACAAAGCTCTAGAGAACAGAAGACCAAGGATAT
GAACCATGAAACTCCGGCCACCACCTTGAACGACGGTCTCCAGACGACGTGGATGGTCG
TGATATGCACCGAGGATCACCAAAAACCTCTCGAATAGACCAAAACAAGAGTACTAATCA
TGAAGAACAACAACCCCTCATGATCAATTACCCTATAGAAAAGCTAGGGTTTCCGTTAG
AGCTAGATCTGATGCCACTACGGTAAATGACGGATGTCAATGGAGAAAATACGGTCAGAA
AATGGCGAAAGGGAATCCATGTCTCGCGCTTATTATCGTTGCACCATGGCCGTTGGATG
TCCTGTCCGTAAACAGGTCAACGATGCGCGGAGGATACAATATCTTGACAACAACGTA
CGAAGGAAACCATAACCATCTCTTCCCCCGTCAGCCACAGCCATGGCTGCAACCACCTC
CGCCGAGCAGCCATGCTCTTATCAGGCTCCTCCTCCAGCAACCTCCACCAAAACACTCTC
TAGCCCCCTCCGCCACGTCATCATCTCTCTACCATAACTTCCCATACACCTCCACAAT
CGCAACACTCTCTGCCTCAGCTCCTTTCCCCACCATAACCTTAGACCTACCAACCCACC
TCGACCGCTACAACCGCCACCGCAGTTTCTAAGCCAGTATGGTCCCGCCGCTTTTTTACC
AAACGCTAATCAAATTAGGTCTATGAATAATAATAACCAGCAGTTATTAATACCTAATTT
GTTTGGCCCAAGCCCCACCGTGAATGGTCGATTGAGTTAGGGCTGCGATTGCGAT
GGATCCGAACCTTACGGCGGCACTTGGCGCCGCGATCTCAAACATTATCGGAGGAGGTAA
TAACGACAACAATAATAACTGATATTAATGATAACAAGGTTGATGCAAAAAGTGGAGG
GAGTAGTAACGGAGATTGCCACAGCTTCTCAGTCTTGCAACACTTTCTCTACAACTA
ATTTTACTACCATTATTATATGTTATCTTATTATATATTACACACATATTATACATTA
TGCGTATCTTAAGTTTTTTTTTGGGGGCCATTATATATGAATGATATGGAGATCACTGAG
AGAGAGAGAGAGCTATTATGGGTTTTTTTTT

>G1417 Amino Acid Sequence (domain in AA coordinates: 239-296)
MEEHIQDRREIAFLHSGEFLHGDSKDHQPNESPVERHHESSIKEVDFFAAKSQPFDLG
HVRTTTIVGSSGFNDGLGLVNSCHGTSSNDGDDKTKTQISRLKLELERLHEENHKLKHL
DEVSESYNDLQRRVLLARQTQVEGLHHKHQEDVPQAGSSQALENRRPKDMNHETPATT
RRSPDDVDGRDMHRGSPKTPRIDQNKSTNHEEQQNPDPYRKARVSVRARS DATTVND
GCQWRKYGQKMAKNPCPRAYRCTMAVGCPVRKQVORCAEDTTILTTTYEGNHNLPLP
SATAMAATTSAAAAMLLSGSSSNLHQLTSSPSATSSSSFYHNFYPYSTIATLSASAPFP
TITLDLTNPPRLPQPPQFLSQYGPAAFLPNANQIRSMNNNNQQLLIPNLFPGPQAPPREM
VDSVRAAIAMPNFTAALAAAIISNIIGGGNNNNNTDINDNKVDAKSGSSNGDSPQLP
QSCCTTFSTN*

>G1442 (1..1293)

ATGGGAACAAGAGCAGAACGCAAGGAAGATTTTGTGGTGGGTTGGATTTGGTGTGTA
GAAATTCGCATAAAGACGTTATGGTGCTACCTCATCATCACTATTATCCATCATATTCA
TCACCTTCTCTTCTTCTTTGTGTTACTGTTCTGCTGGTGTAGCGATCCCATGTTCTCT
GTTTCTAGCAATCAGGCTTACACTTCTTCTCACAGTGGTATGTTACACCCGCCGTTCT
GGTCTGCTGCTGTGACTGTAGCAGATCCTTTTTTCTCCTTGAGCTCTTCAGGGGAAATG
AGAAGAAGTATGAACGAAGATGCTGGTGCAGCTTTCAGCGAAGCTCAATGGCATGAGCTT
GAGAGGCAGAGGAATATATACAAGTACATGATGGCTTCTGTTCTCTTCTCCAGAGCTT
CTCACACCTTTTCCCAAGAACCACCAATCAAACACTAACCCGGATGTAACGTGGCAGTG
GCGACAGGAGGCTCATTCAGCTGGGGATTGCTTCAAGCGCAAGCAATAACACGCTGAT
CTGGAGCCATGGAGGTCAAGAGAACAGATGGGAAGAAATGGAGGTGCTCTAGAAACGTG
ATTCTGATCAGAAATACTGTGAGAGACACACACAAGAGCCGCTCCTCGTTCAAGAAAG
CATGTGGAATCATCTCACCATTATCTCACCACAATGACATTCGTACGGCTAAGAATGAT
ACTAGCCAGCTTGTGAGAACTTATCCTCAGTTTACGGACAACCTATAAGCCAGATCCCT
GTGCTTTCTACTCTTCCGTCTGCTCCTCTCCATATGATCACCACAGAGGACTGAGGTGG
TTTACGAAAGAGATGATGCCATTGGAACCTTAAACCCGGAGACTCAAGAAGCTGTCCAG
CTGAAAGTTGGATCAAGCAGAGAGCTCAAACGGGGATTGATTATGATCTGAATTTTCAGG

CAGAAAGAGCCAAATAGTAGACCAGAGCTTTGGAGCATTGCAGGGTCTATTAAGTCTAAAC
CAGACACCACAACATAACCAAGAAACAAGACAGTTTGTGTAGAGGAAAGCAAGATGAA
GCGATGGGAAGCTCTCTGACACTCTCAATGGCTGGAGGAGGCATGGAGGAAACAGAGGA
ACAAACCAGCATCAGTGGGTTAGCCATGAAGGTCCATCATGGCTCTATTCAACAACACCA
GGTGGACCATTGGCTGAAGCACTGTGTCTCGGTGTCTCCAACAACCAAGTTCTAGTACT
ACTACTAGTAGCTGCAGCAGAAGCTCAAGCTAA

>G1442 Amino Acid Sequence (domain in AA coordinates: 172-223)
MGTRAERKEDFVGFGFVGVVENS HKDVMVLPHHHYPSYSSPSSSLCYCSAGVSDPMFS
VSSNQAYTSSHS GMFTPAGSGSAAVTVADPFFSLSSSGEMRRSMNEDAGAAFSEAQWHEL
ERQRNIYKMMASVPVPELLTPFPKNHQSNTPDVTVA VATGGSLQLGIIASSASNTAD
LEPWRCRKTGKKWRC SRNVI PDQKYCERH THKSRPRSRKHVESSHQSSHHNDIR TAKND
TSQLVRTYPQFYGPISQIPVLSTLPSASSPYDHRGLRWFTKEDDAIGTLNPETQEAVQ
LKVGSSRELKRGFDYDLNFRQKEPIVDQSFGALQGLLSLNQTPQHNQETRQFVVEGKQDE
AMGSSLTSLMAGGMEETEGTNQHQWVSHEGPSWLYSTTPGGPLAEALCLGVSNPNSSST
TTSSCSRSSS*

>G1454 (86..1180)
CTAGTAGTGATGATATGATCGCTTCTTCTCTCTACAATCTCAGAAACCTCCGATCACGGTT
TTAGATATCTTCTACAACGGATAACAATGGAGAGCACCATTCTTCCGGTGGTCCACCACC
GCCACAACCTAACCTTCTCCAGGCTTCCGGTTTACCCTACCGACGAAGAGCTTGTGT
TCACTACCTCAACGCAAGCAGCCTCTGCTCCTTTACCTGTGCGCATCATCGCCGAAGT
CGATCTCTATAAATTTGATCCATGGGAACCTCCCGCTAAAGCATCGTTTGGAGAACAAGA
ATGGTACTTCTTTAGTCCACGAGATCGGAAGTATCCAAACGGAGCAAGACCAACACAGAGC
GGCGACTTCAGGTTATTGGAAAGCGACCGGTACAGATAAACCGGTACTTGTCTCCGACGG
TAACCAAAAGGTGGGCGTGAAGAAGGCACTAGTCTTCTACAGTGGTAAACCAACCAAAAGG
CGTTAAAGTGATTGGATCATGCATGAGTATCGTCTCATCGAAAACAAACCAACAATCG
ACCTCCTGGCTGTGATTTCCGCAACAAAAAACTCACTCAGACTTGATGATTGGGTGTT
ATGTAGAATCTACAAGAAGAACACGCAAGTCGACATGTTGATAACGATAAGGATCATGA
TATGATCGATTACATTTTCAGGAAGATTCCTCCGTCTTTATCAATGGCGGCTGCTTCTAC
AGGACTTCACCAACATCATCATAATGTCTCAAGATCAATGAATTTCTTCCCTGGCAAAAT
CTCCGGTGGTGGTTACGGGATTTTCTCTGACGGTGGTAACAGAGTATATACGACGGCGG
TGGCATGATCAACAATAATTGGTACTGACTCAGTAGATCACGACAATAACGCTGACGTCGT
TGGTTTAAATCATGCTTCGTCGTCAGGTCCTATGATGATGGCGAATTTGAAACGAATCT
CCCGGTGCCGTATTGGCCTGTAGCAGATGAGGAGCAAGATGCATCTCCGAGCAACGGTT
TCACGGTGTAGAGGAGGAGGAGGAGATTGTTTGAACATGCTTCTCCTCATGATGGAAGA
GACTCCACCATTGATGCAACAACAAGGTGGTGTGTTAGGAGATGGATTATTGAAACGAC
ATCGTACCAATTACCCGGTTTAAATTGGTACTCTTCTTAATCAAATGTGTTTCGCCGCCG
GTGTGAAGAATTTTCGGTGACAGTGAAGATTTTTTTCGATTGGTGGGGTCATTTGCAT
GCATTATATAATTTGAGATTTGTGTATATGTTTGGGTTAATTAATTGTCACAGGGGC
>G1454 Amino Acid Sequence (conserved domain in AA coordinates: 9-178)
MESTDSSGGPPPPQPNLPPGFRFHP TDEELVVHYLKRKAASAPLPVAIIAEVDLYKFDPW
ELPAKASFGEQEWYFFSPRDRKYPNGARPNRAATSGYWKATGTDKPVLASDGNQKVG VKK
ALVFYSGKPPKGVKSDWIMHEYRLIENKPNRPPGCD FGNKKNLRLDDWVLCRIYKKN
ASRHVDNDKDHDIMIDYIFRKIPPSLSMAAAS TGLHQHHNVSRSMNFFPGKFSGGGYGIF
SDGNTSIYDGGMINNIGTDSVDHDNNADVGLNHASSSGPMMANLKR TLPPVYPVPA
DEEQDASPSKRFHGVGGGGDCSNMSSSMMEETPPLMQQGGV LGDGLFR TTSYQLPGLN
WYSS*

>G1459 (1..1272)
ATGATGAAAGGTCTGATTGGGTATAGATTTAGTCCGACGGGAGAGGAAGTGATCAACCAT
TACCTAAAGAACAACCTTCTGGGTAAAGTATTGGCTCGTTGATGAAGCTATTAGCGAGATC
AACATCTTGAGTCACAAACCAGCAAGGATTTGCCTAAGTTAGCTAGGATCCAATCGGAA
GATCTTGAATGGTATTTCTTCTCCTCGATTGAGTACACGAACCCGAATAAGATGAAAATG
AAGAGGACGACAGGTTCTGGGTTTGGAAACCTACTGGTGTGATCGGGAAATTAGGGAT
AAAAGAGGAAATGGTGTGTGATAGGGATTAAGAAGACGCTTGTGTACCATGAAGGTAAG
AGTCCTCATGGAGTTAGAACTCTTGGGTTATGCACGAGTATCACATCACTTGCTTGCTT
CATCATAAGAGGAAATATGTTGTCTGCCAAGTAAAGTATAGGGGTGAAGCTGCAGAAAT
TCATATGAGCCAAGTCCCTCTTGGTATCCGATTTCGCATACCGTCATAGCGATTACCGGA
GAACCGGAACCTGAGCTTCAGGTTGAGCAGCCAGGTAAAGAAAATCTCTTGGGTATGTCT

G TAGATGATTTGATAGAACCAATGAACCAACAAGAGGAGCCACAAGGTCCTCACTTAGCT
CCGAATGATGATGAGTTTATACGTGGATTGAGGCATGTTGATCGAGGGACGGTTGAATAT
TTGTTTGCCAATGAAGAAAACATGGATGGTTTGTCTATGAATGACTTGAGAATCCCAATG
ATCGTCCAACAAGAGGATCTCTCTGAGTGGGAGGGATTTAACGCAGACACCTTTTTCAGC
GACAAACAACAATAACTATAACCTTAACGTGCATCATCAACTAACGCCTTACGGCGATGGC
TATTTGAATGCATTTTTCGGGTTATAACGAAGGGAATCCTCCCGATCACGAATTAGTGATG
CAAGAGAACCGCAACGATCACATGCCAAGGAAACCTGTGACAGGGACCATTGATTATAGC
AGCGATAGTGGCAGTGATGCTGGATCCATATCTACAACGGTGAAACAAGAAATCCCAAGA
GCTGTTGATGCACCCATGAACAATGAGTCATCTTTGGTGAAAACAGAGAAGAAAGGCTTG
TTTATTGTAGAGGACGCAATGGAGAGAAACCGCAAGAAACCACGATTTATCTATCTCATG
AAGATGATCATAGGCAACATCATATCGGTTTTACTACCCGTCAAAAGATTGATCCCGGTG
AAGAAGTTATGA

>G1459 Amino Acid Sequence (conserved domain in AA coordinates:10-152)

MMKGLIGYRFSPTGEEVINHYLKNKLLGKYWLVD E A I S E I N I L S H K P S K D L P K L A R I Q S E
D L E W Y F F S P I E Y T N P N K M K M K R T T G S G F W K P T G V D R E I R D K R G N G V V I G I K K T L V Y H E G K
S P H G V R T P W M H E Y H I T C L P H H K R K Y V V C Q V K Y K G E A A E I S Y E P S P S L V S D S H T V I A I T G
E P E P E L Q V E Q P G K E N L L G M S V D D L I E P M N Q Q E P Q G P H L A P N D D E F I R G L R H V D R G T V E Y
L F A N E E N M D G L S M N D L R I P M I V Q Q E D L S E W E G F N A D T F F S D N N N N Y N L N V H H Q L T P Y G D G
Y L N A F S G Y N E G N P P D H E L V M Q E N R N D H M P R K P V T G T I D Y S S D S G S D A G S I S T T V K Q E I P R
A V D A P M N N E S S L V K T E K K G L F I V E D A M E R N R K P R F I Y L M K M I I G N I I S V L L P V K R L I P V
K K L *

>G1460 (87..995)

C G T C G A C C T T C A C T C A A A C C C T A A T C C C G G G A A C C C G G G A A T T T T G A T C A T T T T G T T T C T
T T T C G A T C T G T T T C A T T T T A A A A A G A T G A T G A A A G A T C C G A C T G G G T A T A G A T T T A G T C
C G A C G G G A G A G G A A G T G A T A A A C C A T T A C C T A A A G A A C A A A A T T C T G G G T A A G A C T T G G C
T C G T T G A T G A A G C C A T T A G C G A G A T C A A C A T C T T G A A T C A C A A C C C A G C A A G G A T T T G C
C T A A G T T A G C T A G G A T C C A A T C G G A A G A T C T T G A G T G G T A C T T T T T C T C C G A T T G A G T
A C A C G A A C C C G A A T A A G A T G A A A A T G A A G A G A C G A C A G G T T C T G G G T T T T G G A A C C T A
G T G G T G T T G A T C G G A A A T T A G G G A T A A A G A G G A A A T G G T G T T G T G A T A G G G A T T A A G A
A G A C G C T T G T G T A C C A T G A A G G T A A G A G T C C T C A T G G A G T T A G A A C T C C T T G G G T T A T G C
A C G A G T A T C A C A T C T T G C T T G C C T C A T A T A A G A G G A A A T A T G T T G T C T G C C A A G T A A
A G T A T A A G G G T G A A G C T G C A G A A A T T C A T A T G A G C C A A G T C C C T C T T T G G T A T C C G A T T
C G C A T A C C G T C A T A G C G A T T A A C G G A G A A C C G A A C C T G A G C T T C A G G T T G A G C A G C C A G
G T A A A G A A A A T C T C T T G G G T A T G T C T G T A G A T G A T T T G A T A G A A C C A A T G A A C C A A C A A G
A G G A G C C A C A A G G T C C T C A C T T A G C T C C G A A T G A T G A T G A G T T T A T A C G T G G A T T G A G A C
A T G T T G A T C G A G A G C C G T T G A A T A T T T G T T T G C C A A T G A A G A A A C A T G G A T G G T T T G T
C T A T T A T G A A T G A C T T G A C A A T C C C A A T G A T C G C C C A A C A A G A G G A T C T C A T T C T C T G
A G T G G G A G G G A T T A T C G C A G C C A C C T T T T C A G C G A C A A C A A C A A T A A C A A T A A C C T T A
A C G T G C A T C A A C T A C G T C T T T C T T A C C G G A T G A T T A T C A G A A T G C A T T T T G G G T T A C A
A C G G A G C G N C C G C T

>G1460 Amino Acid Sequence (domain in AA coordinates: TBD)

M M K D P T G Y R F S P T G E E V I N H Y L K N K I L G K T W L V D E A I S E I N I L N H K P S K D L P K L A R I Q S E
D L E W Y F F S P I E Y T N P N K M K M K R T T G S G F W K P S G V D R K I R D K R G N G V V I G I K K T L V Y H E G K
S P H G V R T P W M H E Y H I T C L P H H K R K Y V V C Q V K Y K G E A A E I S Y E P S P S L V S D S H T V I A I N G
E P E P E L Q V E Q P G K E N L L G M S V D D L I E P M N Q Q E P Q G P H L A P N D D E F I R G L R H V D R E P V E Y
L F A N E E N M D G L S I M N D L T I P M I A Q Q E D L I L S E W E G F I A A T F F S D N N N N N N L N V H Q L T S F L
P G *

>G147 (37..672)

A A A T C A T C A G A T A G A A G G A A A T A T T C T G A T T G A G A G A T G G C T C G T G G A A G A T T C A G C T T
A A G A G G A T T G A G A A C C G G T T C A C A G A C A A G T G A C T T T T T G C A A G A G G A G A A C T G G T C T T
C T C A A G A A G G C T A A G G A G C T C T C T G T G C T C T G T A T G C C G A G A T C G G T G T T G T G A T C T T C
T C T C C T C A G G G C A A G C T C T T G A G C T C G C T A C T A A A G G A A C A A T G G A G G G A A T G A T T G A T
A A G T A C A T G A A G T G A C T G G T G G T G G T C G T G G T C T T C T T C T G C T A C T T T T A C T G C T C A A
G A C A A C T T C A A C C A C C A A A T C T T G A T C C G A A A G A T G A G A T C A A C G T G C T T A A G C A A G A G
A T T G A G A T G C T T C A G A A A G G G A T A A G C T A T A T G T T T G G A G G A G G A T G G G G C T A T G A A T
C T T G A A G A A C T T C T T T T G C T T G A G A A G C A T C T T G A G A T T G G A T T T C T C A G A T T C G C T C T
G C T A A G A T G G A T G T T A T G C T T C A A G A A T T C A G T C A T T G A G G A A C A A G G A A G G A G T C C T C

AAAAACACCAACAAGTATCTCCTCGACAAGATAGAGGAAAACAACAATAGCATATTAGAT
GCTAACTTCGCAGTCATGGAGACAACTATTCCTATCCGCTAACAAATGCCAAGTGAAATA
TTTCAGTTCTAGACCATAGGGTATTTGAAGACTATGTCTCACGAATTTAAATAACCTTGG
TAAGTATAATATAGTGTGTAAATCACACATAATTAATAAAGCCTGTGGAACCTTCGC
TAGGCAGTTGAAAATCTATCCGTATGTTTTATCCTCTTGTGTTTACATTTGTTGGTGTGAA
GATGAAATGACTGCAAGTGTGGTGTGTACTTATAACTCTTCTACTTTCTATCTATGTTT
TGAATTTATGGATT

>G147 Amino Acid Sequence (domain in AA coordinates: 2-57)
MARGKIQLKRIENPVHRQVTFCKRRTGLLKKAKELSVLCDAEIGVVIFSPQGKLFELATK
GTMEGMIDKYMKTCTGGGRGSSSATFTAQEQLPNLDPKDEINVLKQEIEMLQKGISYMF
GGGDGAMNLEELLLLEKHLEYWISQIRSAKMDVMLQEIQSLRNKEGVLKNTNKYLLDKIE
ENNNSILDANFAVMEYNYSYPLTMPSEIFQF*

>G1471 (1..735)
ATGGAGAACCAATCTATGTCTTCATCAAGCTCCTCCACACACAAACATGATCAAAAACCTC
AAAAGTTCCGTTGTGGCCATGGAGGTCCTGGAGGAGAAGGAGACAGTGAACAATCCGCCC
CAGTATTATAATAAGATCTACATCTGTTACTTGTGCAAGAGAGCGTTCCCAACCCCTCAT
GCCCTTGGCGGTACGGAACCAACCCACAAGGAGGACCGAGAATTGGAGAGGCAACAGATC
GAGTCAAGGCTTTCTAACAAAGACAAGTCTAACTTGCTCTTGGTGGGTCTTCACAAGAT
GTTTTATCAAATGATAATCACCTTGGACTCTCTCTTGGTCCATGAAGTCCATAGAAGGT
AGCAGCAGCAGCAACAACGTTAACCCATTGCTTAATGTTGGAGTCCCTAGAGGAACCACA
GATATGAACATGAACAACCTATAGCTCACATGCTTTATCAACTGATGATATTAATCTTGAT
CTTACTCTTGGTCCATCTAAGTCCATAGGAGATAGCAACAATATCATTAAATAACAACACT
AACTCATCCTTCGATGGGAATCTGATCATTTCCCGTTCTCGTGTGTCTAGATACCAT
TTTGTGCTGGGAACCCCTTGATTCAATCTCTAGAAACATTCCCTCCTTCTATTACTTTT
CCTCATCTAAACATCAATCTTTCTCATGATTCGTTTTCTTTACAAGAGAATGGTTCGGGC
TCTAGTCACTCATAA

>G1471 Amino Acid Sequence (domain in AA coordinates: 49-70)
MENQSMSSSSSTHKHDQKLKSSVVAMEVLEEKETVNNPPQYNNKIYICYLCKRAFPPTPH
ALGGHGTTHKEDRELERQOIESRLSNKDKSNLLFGGSSQDVLSDNHLGLSLGPLKSI
SSSSNNVNPLLNVGVPRTTDMNMNYS
SHALSTDDINLDTLGLPSKSI
GDSNNIINNNT
NSSFDGNLIIPVRPRVSRVYHFVAGNPLDSISRNP
PSITFPHLNINLSHDSFSLQENGSG
SSHS*

>G1475 (1..645)
ATGAAGAGAACACATTTGGCAAGTTTGTAGTAACAGAGACAAAACCAAGAAGAAGAAGGA
GAAGACGGTAATGGTGACAACAGAGTCATCATGAATCACTACAAGAATTACGAAGCTGGG
CTGATCCCATGGCCTCCCAAGAATTACACTTGCAGCTTCTGCAGGAGAGAGTTCAGATCT
GCTCAAGCACTTGGAGGCCACATGAATGTTCATAGAAGAGACAGAGCAAAACTCAGGCAG
ATCCCTTCTTGGCTCTTCGAACCTCACCACCACACACCTATTGCAAACCTTAACCTAAT
TTTAGCTCTTCTTCTTCTTCTTCAACAACAACAGCTCATCTTGAGCCTTCCCTAACCAAC
CAGAGATCCAAAACCACTCTTTTCTCTTCTGCCCCGTTTGATCTTTTGGACAGTACTACT
AGCTATGGAGGTTTGATGATGGACAGAGAGAAGAACAAGAGCAATGTATGTAGCAGAGAG
ATCAAGAAAAGTGCCATCGATGCATGTCATTAGTAAGATGTGAGATAAGCCGTGGGGAT
CTGATGAATAAGAAAGATGATCAAGTCATGGGGTTGGAGCTTGGGATGAGTTTGAGGAAT
CCCAACCAAGTTCTTGATTGGAGCTTCGACTAGGCTACCTCTAA

>G1475 Amino Acid Sequence (domain in AA coordinates: 51-73)
MKRTHLASFSNRDKTQEEGEDGNGDNRVIMNHYKNYEAGLIPWPPKNYTCSPCRREFRS
AQALGGHMNVHRRDRAKLRQIPSWLFEPPHHHTPIANPNPNFSSSSSSSTTAHLEPSLTN
QRSKTTFPFSARFDLDDSTTSYGGLMMDREKNKSNVCSREIKKSAIDACHSVRCEISRGD
LMNKKDDQVMGLELGMSLRNPQVLDLELRLGYL*

>G1477 (1..606)
ATGTTGTCTCTCGACTCGAATTACGCTAGTGATATTAGCGACGATGCCTCCGCCACCGGA
TCGATAGAGAATCCTATATACAAATGCAAGTATTGTCTAGGAAGTTCGATAAAACACAA
GCATTAGGTGGTCATCAAAATGCACACAGAAAGGAGAGAGGTTCGAAAAACAACAAAAA
GCATTTTGGCGCATTTGAACCGACCAGAACCAGATCTTTACGCGTACTCGTATTTCGTAT
CATCATTCATTTCTAACCAATACGCACTCCACCCGGGATTTGAACAGCCTCAGTACAAA
GTTGATAGATCATACAAGATGTCCATGGTCTACAACCAATATGTGGGATCCTCAAGCTCT
AGCTTTGCAGGACTACAAAGTGACCCAAGTCAAGGAATGAACCAGGATTGGACCTTTACC

GGGATCCCATTCCTACCCCAATCTCAACCTCAACCCTATCGTCACCAATA'GTTTGGAT
CTTTGCCTTGGCATTGGTAGCTCCCAAACCCAAACCACAACCTCAAGAACCAAAATGATGCA
ACAGAAGAGATGGATGCTGAGAAAGAAAATGATGGTTCTTCCCTTTCTCTCTCACTCAAA
CTGTGA

>G1477 Amino Acid Sequence (domain in AA coordinates: 29-48)
MLSSDSNYASDISDDASATGSIENPIYKCYCPKFDKTQALGGHQNAHRKEREVEKQQK
AFLAHLNRPEPDLAYSYSYHHSFPNQYALPPGFEPQYKVDRSYKMSMVYNQYVGSSSS
SFAGLQSDPSQGMNQDWTFTGIPFLPQSQPQLSSPICLDLCLGIGSSQTQPQPQEPNDA
TEEMDAEKENDGSSLSLSLKL*

>G1487 (1..1020)

ATGGAACAAGCCCGCTTGAAGAGCAGCGTCAGGAAAGAGATGGCTCTCAAAACGACTTCT
CCGGTTTACGAAGAGTTTCTTGCCGTCACCACCGCTCAAAATGGCTTTTCCGTCGACGAT
TTCTCTGTAGACGACTTGCTTGACTTGTCAAACGATGACGTTTGTGCGACGAAGAACT
GACCTCAAGGCTCAACATGAGATGGTCCGTGTTTCTCTGAGGAACCAACGACGACGGA
GACGCTCTTCGCGGAGCAGCGATTTCTCCGGCTGTGACGACTTTGGTTCTCTCCCTACA
AGCGAACTCTCTCTTCGCGGATGATTTAGCGAACCTTGAGTGGCTCTCTCATTTCTGTG
GAGGACTCCTTCACGGAATATTCGGGTCCAAACCTCACCGGAACCCGACTGAGAAACCG
GCGTGGTTAACGGGTGACCGGAAACATCTGTGACTGCAGTCACGGAAGAGACCTGTTTC
AAATCCCCGTGTTCGGGCTAAAGCCCGTAGCAAACGTAACCGCAATGGCCTCAAGGTCTGG
TCGCTTGGTTTCGTCGCTCTCTCGGGTCTTCTCGTCCGGTTCGACCTCCTCCTCTCT
TCGGGTCTTCCAGCCCGTGGTTCTCCGGCGCTGAGCTGCTCGAGCCTGTGGTCACGTCA
GAGAGGCCACCGTTTCCCAAGAAGCATAAGAAAAGGTGAGCCGAGTCTGTTTTCTCCGGT
GAGCTGCAGCAGCTGCAACCTCAGCGAAAGTGACGCACTGCGGCGTTTCAAGAACTCCG
CAGTGGAGAGCCGGGCAATGGGAGCCAAGACCCTGTGCAATGCGTGCAGGTGTCCGGTAC
AAGTCCGGTAGGTTGCTACCGGAATACAGACCCGCTTGAGTCCCGACATTCTCGAGTGAG
CTGCACTCGAACCACCGGAAAGTCATAGAGATGAGGCGGAAGAAGGAGCCAACAGT
GACAACGAAACCGGTTTAAACCAGCTGGTTCAGTCCCAAGCTGTACCAAGTTTTTGA

>G1487 Amino Acid Sequence (domain in AA coordinates: 251-276).
MEQAALKSSVRKEMALKTTSPVYEEFLAVTTAQNFGSVDDFVDDLLDLSNDDVFADDEET
DLKAQHEMVRVSSEEPNDDGDALRRSSDFSGCDDFGSLPTSELSLPADDLANLEWLSHFV
EDSFTEYSGPNLTGTPTEKPAWLTGDRKHPVTAVTEETCFKSPVPAKARSKRNRNGLKVV
SLGSSSSSGPSSSGSTSSSSSGPSSPWFSGAELLEPPVTSERPPFPKHKHRSASVFSG
ELQQLQPQRKCSHCGVQKTPQWRAGPMGAKTLCNACGVRYKSGRLLPEYRPACSPTFSSE
LHSNHRKVIEMRRKKEPTSDNETGLNQLVQSPQAVPSF*

>G1492 (149..919)

AATCCCAACCCACACACCTCTCAAATCCTCCTCTCCTCGTTTCTCTTTCTCTCCTCTTCA
CAGAACCAAAACATATCAAACCTTTTTTCTCTTGGGTTTAAAGTAAAAATCGAATCTTTG
TGTCGGTTTTTAGGGTTCTTGAAACGATATGGGTAAGTCTAGTGGTAGAAATGGTAACGG
AAGCTTTAACGGCAATAAATTTACGGAGTTAGACCTTACGTACGGTCTCCAGTTCCACG
GCTTAGATGACGACCGCGGATCTTACCGTTGTTTTCGTTACGCCGTCGAGATTCTCGGTGG
TCAACACCGAGCAACACCAAACTTGTCTTAAAGATGATGGATGTGAAGGGACTTACCAT
TTCACATGTCAAAGCCACCTTCAGATGTATAGAGGAGGTTCAAAGCTCACTTTGGAGAA
ACCAGAAGAAAGCTCATCATCTTCAATAAGAAGAAGACAAGACAGTGAAGAAGATTATTA
TCTTCATGACAACTTGTCTTTACACACAAGGAATGATTGTCTTTTGGGTTTTCACTCTTT
TCCTCTTTCTTACATTCTTCATTTAGAGGAGGAGGAGGAGGAAGAACAAGAGCAGCA
GACTTCAGAGTCTGGTGGTTATGATGATGATGCTGACTTCTTACATCAAGAAGATGAA
CGATACGACGACGTTTTTGTTCACATCATTTCCCAAGGGAACAGAGGAGTGGCGGGAACA
AGAACACGAAGAAGAAGAAGATTTGTCTCGTTGTCTCTGTCTGTTAAATCATCATATTG
GAGAAGCAATGGATCATCGGTGGTGAGCGAAGAGTGAAGCAGCAGTCTCGACTTGTTC
AGCACCATTTCGTATCCAAAGATTGCTTTGGTTCTTCAAAGATTGATCTTAATCTGTCAAT
TTCTCTCTCGGTAGCTAAATAAGTTATGCAAGATTTAGGTTTCAAGAACTATTTCGGAT
GTGTTTTTGAAGTATGATGTTAGTAGAGAAACCTAGAAAATGAAGTTTAGAT
AAATTATCAACGACGCGTTTTGATCGCCTTTGAACGGAATAAATAA

>G1492 Amino Acid Sequence (domain in AA coordinates: 34-83)
MGKSSGRNGSGFNGNKFHGVPRPVPRPRLRWTPDLHRCFVHAVEILGGQHRATPKLV
LKMDVKGLTISHVKSHLQMYRGGSKLTLEKPESSSSSIRRRQDSEEDYYLHDNLSLHT
RNDCLLGFHSFPLSSHSSFRGGGGGRTKEQQTSESGGYDDADFLHIKKMNDTTTFLSHH

ATGCTCTCAGTCAAACATGGTACCAGTGGCTAACAAACGGAGACAACAACAACGCAACAGGAA
AACAACAACAACAACAACAATGGTGGAACTGACAACACTAATGCTGGAAATGATTCT
GGAGATCAAGATTTTCGACAGTGGGAATACCTCAAGTGGCAATCATGGAGAAGGGTTGGGA
ACAATCAAGCTCCTCGTCATAAGAAGAAAAAATACAATCGTCACACCCAACTTCAGATT
TCGGAGATGGAAGCTTTCTTCAGAGAGTGTCCTCACCCAGATGACAAACAAGGTACGAC
CTTAGCGCTCAATTGGGATTTGACACCTGTTTCAGATCAAAATCTGGTTCCAGAAACAACG
ACTCAAAAACAAGATCAACAAGAACGCTTTTGAGAACTCAGAACCTCGGAATCTGAACAAC
CACCTTAGGTCTGAAAATCAGCGGTTACGAGAAGCTATTCTACAAAGCCTTATCGCCCTAAG

TGTGGAGGCCAACTGCAATTGGCGAAATGACCTTCGAAGAGCACCATCTTCGCATCCTC
AACGCTCGTTTTGACTGAAGAGATCAAGCAACTTTCCGTGACAGCGGAAAAGATATCAAGG
CTTACGGGGATACCAGTAAGGAGCCATCCCCGTGTGTCTCTCTAATCCTCCTCCAAAT
TTCGAGTTCCGGATGGGATCTAAGGAAAATGTCGGAAACCACTCGAGGGAAACCACTGGA
CCTGCAGATGCTAATAACCAAGCCGATCATCATGGAGTTGGCATTGGAGCCATGGAGGAG
CTCTTGGTGTAGGCTCAAGTGGCTGAACCACTGTGGATGGGAGGATTAAATGGCACTAGC
TTAGCTTTGAACTTGGATGAATACGAAAAGACGTTTCGCACGGGTCTCGGTCTAGACTT
GGCGGGTTTCGAACCGAGGCATCCAGGGAACTGCACTCGTGGCAATGTGTCTACTGGC
ATTGTTGAAATGCTCATGCAAGAGAATCTGTGGTCAACAATGTTTGGCGGAATTGTTGGT
AGAGCCAGGACTCATGAACAGATAATGGCTGATGCTGTGGAACTTCAATGGAAATCTC
CAAATAATGAGTGTGAGTACCAAGTGTCTTCCCCGCTAGTCACAACCCGCGAAAGCTAC
TTCGTCCGCTACTGTAAGCAACAAGGAGAGGGTTTGTGGGCGGTGGTTCGATATTTCCATC
GACCATCTCTCCCAAACATCAACCTAAAAATGTCGCCGCCGACCCTCTGGATGTCTGATT
CAAGAAATGCATAGTGGTTACTCCAAGGTTACATGGGTGGAACATGTGGAAGTAGATGAT
GCAGGAAGTTACAGCATCTTTGAGAAATTAATCTGTACTGGTCAAGCTTTTGTCTGCTAAC
CGCTGGGTTGGTACATTGGTACGCCAGTGTGAGCGGATATCTAGCATCTTGTGACAGAT
TTTCAATCTGTGATTCCGGTGTATCACATAACGCTAACTAACCATGGAAAGATGAGCATG
CTGAAGATAGCTGAGCGGATTGCGAGAACCTTCTTTGCTGGAATGACCAATGCGACGGGG
TCTACAATATTTTCTGGTGTGAAGGAGAAGATATCAGAGTGTGACAATGAAGAGCGTG
AATGATCCAGGAAAGCCTCCCGGTGTCTATTATTTGTGACGCCACTTCTTTTGGCTTCTC
GCTCCTCCTAACACTGTCTTTGACTTCTCTCAGAGAGGCTACTACCGACACAATTGGGAT
GTTCTCTGCAACGAGAGATGATGCACAAGATAGCAGAGATTACGAATGGGATAGACAAA
AGGAAGTGTGCAAGTTTACTCCGGCATGGACACACTAGCAAGAGCAAGATGATGATAGTT
CAAGAGACTTCTACTGACCAACAGCTTCATTTGTGCTTTATGCGCCTGTTGATATGACA
TCAATGGATATTACTCTCCATGGAGGTGGTGTCTGACTTTGTGGTGATCCTGCCTTCT
GGTTTGTCTATTTTCCAGATGGTACGGGTAAGCCTGGAGGAAAAGAAGGAGGATCACTT
TTGACCAATTCTCTTCCAAATGCTGGTTGAGTCAAGTCTGAGGCTAGGCTGAGTGTAGC
TCTGTTGCAACTACTGAGAATCTGATTCTGTACAACCGTGGGAGGATCAAAGATTTGTTT
CCTTGTGCACTGCTTGA

>G1544 Amino Acid Sequence (domain in AA coordinates: 64-124)
MSQSNMVPVANNGDNNNDNENNNNNNNNGGTDNTNAGNDSGDQDFDSGNTSSGNHGEGLG
NNQAPRHKKKKYNRHTQLQISEMEAFFRECPHPDDKORYDLAQGLDLPVQIKFWFQNKR
TONKNQOERFENSELRNLNNHLRSENQRLREAIHQALCPKCGGQTAIGEMTFEEHHLRIL
NARLTETIKQLSVTAEKISRLTGI PVRSHPRVSPNPPNPFEGMGSKGNVGNHSRETTG
PADANTKPI IMELAFGAMEELLVMAQVAEPLWMGGFNGTSLALNLDYEYKTFRTGLGPRL
GGFRTEASRETALVAMCPTGIVEMLMQENLWSTMFAGIVGRARTHEQIMADAAGNFNGNL
QIMSAEYQVLSPLVTTRESYFVRYCKQQEGELWAVVDISIDHLLPNINLKRRRPSGCLI
QEMHSGYSKVTWVEHVEVDAGSYIFEKLI CTGQAFANRWVGT LVRQCERISSILSTD
FQSVDSGDHITLTHNGKMSMLKIAERIARTFFAGMTNATGSTIFSGVEGEDIRVMTMKSV
NDPQKPPGVII CAATSFWL PAPPNTVFDLREATHRHNDVLCNGEMMHKIAEITNGIDK
RNCASLLRHGHTSKSKMMIVQETSTDPTASFVLYAPVDMTSMITLHGGGDPDFVILPS
GFAIFPDGTGKPGGKEGSLTISFQMLVESGPEARLSVSSVATTENLIRTTVRRIKDLF
PCQTA*

>G156 (39..755)
AGGAAGAGGGAGCCACTCATAAGAGGAAGAAGAGAGAGATGGGTAGAGGGAAGATAGAGA
TAAAGAAGATAGAGAATCAGACGGCGAGGCAAGTGACCTTCTCCAAGAGAAGAACTGGTC
TTATAAAGAAGACTCGTGAGCTCTCTATTCTCTGTGACGCTCACATCGGTCTCATCGTCT
TCTCAGCCACCGGAAAGCTTTCAGATTCTGCTCCGAACAGAACAGGATGCCTCAACTCA
TTGACCGATACCTGCATACCAACGGATTGCGACTTCTGATCATCATGACGACCAGGAGC
AATTGCACCATGAGATGGAACCTACTAAGAAGAGAGACATGTAACCTTGAGCTTCGTCTGC
GTCCATTCCATGGACATGACTTAGCCTCATTCTCTCTAATGAGCTTGACGGACTCGAGA
GACAGCTAGAACATTCTGTCTCAAAGTCCGTGAGCGTAAGAGGAGGATGCTAGAAGAAG
ATAACAACAACATGTACCGTTGGCTTCATGAGCATCGTGCAGCGATGGAGTTTCAACAAG
CTGGGATAGATACCAAAACAGGGAGTATCAACAGTTTATAGAGCAGCTTCAGTGCTATA
AACCAGGGGAGTATCAGCAGTTTCTAGAGCAGCAGCAACAACAACCAACAGCGTTCTTC
AGCTTGCTACACTTCTTCTGAGATTGATCCTACTTACAATCTCCAGCTTGCTCAGCCTA
ATCTTCAAAACGATCCAACGGCCAGAAATGATTAATACAATTCTCAATAGATATCTACTC

84

TGTTACTACTACGAGTTCATGCCTCTGAAGAACTGA
 >G1587 Amino Acid Sequence (conserved domain in AA coordinates: 61-121)
 MGYISNNNLINYLPLSTTQPPLLLTHCDINGNDHHQLITASSGEHDIDERKNNIPAAATL
 RWNPTPEQITTTLEELYRSGTRPTTEQIQIASKLRKYGRIEGKNVFWFQNHKARERLK
 RRRREGGAIKPKHKVDKSSSGHRVDQTKLCPSPHTNRPQPQHELDPASYNKDNANN
 EDHGTTEESDQRASEVGKYATWRNLVWTSITQQPEEINIDENVNGEEETRDNRTLNLFP
 VREYQEKTRGLIEKTKACNYCYYYEFMPLKN*
 >G1588 (1..2232)

ATGTACCATCCAAACATGTTTGAGAGCCATCATATGTTTCGATATGACCCCAAAGAGTACC
 TCTGATAACGACTTGGGAATCACCGGTAGCCGAGAAGATGACTTTGAGACCAAGTCAGGT
 ACCGAAGTCACTACTGAGAATCCTTCTGGTGAAGAGCTTCAAGATCCTAGCCAACGTCCC
 AACAAAAGAAGCGTTACCATCGCCACACGCAACGCCAAATTCAAGAGCTCGAATCATTC
 TTTAAGGAATGTCCTCATCCAGATGATAAGCAACGAAAAGAGTTGAGCCGTGATCTCAAT
 TTAGAGCCTCTTCAAGTTAAGTTTGGTTCCAAAACAACGCACACAGATGAAGGCACAA
 AGTGAGAGGCATGAGAACCAGATTCTAAAGTCAGACAATGACAAGCTCAGAGCAGAGAAC
 AATAGATACAAAGAAGCTCTAAGCAATGCTACATGCCCTAACTGTGGCGGTCCAGCTGCT
 ATTGGAGAAATGCTTTTGACGAACAACATCTCAGGATCGAAAATGCTCGGCTCCGCGAA
 GAGATTGATAGGATCTCTGCTATTGCTGCGAAATACGTTGGGAAGCCGTTAGGATCGTCT
 TTCGCTCCACTAGCGATCCACGCGCTTCTCGTTTCGTTGATCTTGAAGTTGGAACTTT
 GGGAAACCAGACAGGCTTTGTAGGAGAAATGTATGGAACAGGGGACATTTTGAGGTCACTT
 TCGATTCTTCTGAGACTGATAAGCCTATAATCGTGGAGCTAGCGGTTGCAGCTATGGAG
 GAACTCGTGAGAATGGCTCAAACCTGGAGATCCTTTATGGCTTCAACCGATAATTCAGTC
 GAGATTCTCAACGAAGAAGAGTATTTCAAGACGTTTCCGAGAGGAATTGGACCAAAGCCA
 TTAGGATTAAAGATCAGAGGCGTCAAGACAATCTGCAGTTGTTATAATGAATCACATCAAT
 CTCGTTGAGATTCTCATGGATGTGAATCAATGGTCTTGTGTTTTCTCTGGGATTGTGTCA
 AGAGCCTTGACACTTGAAGTTCTTTCAACTGGAGTTGCTGGGAACACAACGGTGCTTTA
 CAAGTGATGACAGCTGAGTTTCAAGTTCCATCACCCCTAGTCCCAACGCGTGAGAACTAC
 TTTGTGAGATACTGCAAACAACACAGTGACGGCTCTTGGGCTGTGGTTGATGTCTCTTTG
 GACAGCCTTAGACCAAGTACTCCAATCTTAAGAACTAGAAGAAGGCCTTCAGGTTGTCTG
 ATTCAAGAATTGCCTAATGGTTATTCTAAGGTTACATGGATAGAGCATATGGAGGTAGAT
 GATAGATCAGTTTCAACATGTATAAACCGTTGGTTTCACTCCGTTTAGCTTTTCGGTGCG
 AAACGTTGGGTGGCTACACTCGAACGACAATGCGAGCGGCTTGCTAGCTCCATGGCCAGC
 AACATTCTGGTGATCTTTCCGTGATAACGAGTCTGAAGGAAGGAAGAGTATGTTGAAG
 CTAGCTGAGAGAATGGTTATGAGTTTCTGCAGTGGTGTGGCGCGTCACTGCACACGCT
 TGGACAACAATGTGACAACAGGATCCGATGATGTTCCGGTCAAGAGTATG
 GATGATCCAGGAAGACCTCCGGGTATTGTTCTTAGTGACGCTACTTCATTCTGGATCCCA
 GTTGCTCCCAAACGTGTTTTTGTATTTCTCCGTGACGAAAATTCAAGAAAAGAGTGGGAT
 ATTCTGTCAAATGGAGGTATGGTTTCAAGAAATGGCTCATATAGCCAATGGTCATGAACCT
 GGAAACTGTGTCTCCTTGTCTCCGAGTCAATAGTGGAACTCGAGCCAGAGCAACATGTTG
 ATTCTACAAGAGAGCTGTACAGATGCATCAGGATCGTATGTGATTACGCGCCAGTGGAT
 ATAGTGGCGATGAATGTGGTTCTAAGCGGTGGAGATCCTGATTACGTGGCGTTGTTGCCG
 TCTGGTTTTGCTATTTTACCGGATGGTTCCGTTGGAGGAGGAGATGGGAATCAGCATCAG
 GAAATGGTTTTCTACTACTTCTTCTGGGAGTTGTGGTGGTTTCGTTTTAACCGTTGCGTTT
 CAGATTCTTGTGACTCTGTTCTTACAGCTAAACTCTCACTTGGCTCGGTGGCTACGGTT
 AATAGTCTGATCAAATGTACGGTGGAGAGGATTAAAGCTGCTGTTTCTTGTGATGTTGGA
 GGAGGAGCGTAG

>G1588 Amino Acid Sequence (domain in AA coordinates: 66-124)
 MYHPNMFESHMFDMPKSTSDNDLGITGSREDDFETKSGTEVTTENPSGEELQDPSQRP
 NKKRYHRHTQRQIQELSFKECPHPDDKQKELSRDLNLEPLQVKFWFQNKRTQMKAO
 SERHENQILKSDNDKLRANENRYKEALSATCPNCGGPAAIGEMSFDEQHLRIENARLRE
 EIDRISAIAAKYVGKPLGSSFAPLAIHAPSRSLDLEVGNGFNQTFVGMVGTGDIILRSV
 SIPSETDKPIIIVELAVAAMEELVRMAQTGDPLWLSTDNSVEILNEEYFRFFRGIGPKP
 LGLRSEASRQSAVVIMNHINLVEILMDVNQWSCVFSGIVSRALTLEVLSTGVAGNYNGAL
 QVMTAEFQVPSPLVPTRENYFVRYCKQHS DGSWAVVDVSLDSLRLPSTPIILRTRRRPSGCL
 IQELPNYSKVTWIEHMEVDDRSVHNMYKPLVQSGLAFAKRWVATLERQ CERLASSMAS
 NIPGDL SVITSPEGRKSMKLAE RMVMSFCSGVGASTAHAWTTMSTTGSDDV RVMT RKS M
 DDPGRPPGIVLSAATSFWIPVAPKRVDFDLRDENSRKEWDILSNGGMVQEMAHIANGHEP

GNCVSLLRVNSGNSSQSNMLILQESCTDASGSYVIYAPVDIVAMNVVLSGGDPDYVALLP
SGFAILPDGSGVGGDGNQHQMSTSSGSCGSSLTVAFAQILVDSVPTAKLSLGSVATV
NSLIKCTVERIKAASCDVGGGA*

>G1589 (179..2221)

ACCAAACCTACATAGCAATCACACACATCTCCACAAACACAGCTTGAGATGATCATGAAA
CACGTGCATCCTCAGATCTCTATCAATCCAGCTTGGTGAAAGAAGGTCAAGAATTGAAAG
AGAATCAAAGAAAACGACGTCGTTTCATTTCGTGTGTAACTACTAATTATACATAGAT
GGCTGCTTACTTTACGGAAACCCACCGGAGATCTCTGCCGGATCCGACGGTGGTCTTCA
AACGTTGATCCTCATGAATCCAATACTTACGTTTACGTACACCCAACAAGACAACGACTC
GAACAACAACAACAACAGCAACAATAGCAACAACAACAACAACAACAACAACAACA
CAACAACAGTAGTTTCGTTTTCCTCGATTCCACGCGCCGACGCAAAACGCGAGCCAGCA
GTTTCGTCGGAATACCACTCTCAGGTACGAAGCTGCTTCCATTACAGCCGCCGACAACAT
CTCCGTACTTCACGGTTATCTCCGCGCGTGCAGTACAGTCTTACGGTAGCCACCAAGT
GGATCCCACTCACCAGCAAGCCGCGTGTGAGACGCCACGCGCGCAGCAAGGCCTCTCTTT
AACCCTCTCGTCTCAACAGCAGCAGCAACAGCAACATCATCAACAACACAGCCTATTCA
CGTCGGATTCCGGGTCGGACATGGAGAAGATATCCGGGTCTGGCTCTACAGGATC
GGGGGTAACAACAACGGTATAGCTAATCTTGTAGCTCCAAGTACTTGAAGGCAGCACAAGA
GCTTCTTGACGAAGTAGTCAACGCTGATTCCGATGACATGAACGCTAAATCCCACTATT
CTCATCGAAAAAGGGTAGTTGCGGAAATGATAAACCCTGTCGGAGAATCATCGGCCGGCGC
TGGAGGAGAAGGTTCCGGTGGCGGAGCAGAAGCAGCCGGGAAACGTCCGGTGGAGCTAGG
CACGGCAGAGAGACAAGAAATACAGATGAAGAAAGCAAACTTAGTAACATGCTTCATGA
GGTGGAGCAGAGATATAGACAGTACCACCAGCAGATGCAGATGGTGATCTCTTCGTTCTGA
GCAAGCGGCAGGGATAGGATCAGCGAAGTCATACACGTCGCTAGCATTGAAAACCATATC
AAGACAGTTCGCTTGCTTGAAAGAGGCGATCGCTGGTCAGATAAAAGCGGCCAACAAAGAG
TCTTGGGGAGGAAGATTTCAGTGCTTGGTGTGGGAGGTTTGAAGGTCGAGGCTCAAGTT
CGTGGACCACCACTTGAGACAGCAAGAGCTCTTCAACAACGGGAATGATTCAACATCC
TTCCAATAATGCTTGAGACCTCAACGTGGTCTCCAGAACGAGCCGTCTCAGTTCTCCG
TGCTTGGCTCTTCGAACACTTTCTTCATCCATACCCTAAGGATTCCGACAAGCACATGCT
AGCTAAGCAAACAGGACTCACTCGTAGCCAGGTGTGCAACTGGTTTATAAACGCGAGAGT
TCGGTTATGGAACCAATGGTGGAGGAGATGTACATGGAGGAATGAAGGAGCAGGCAAA
GAACATGGGATCCATGGAAAAGACTCCTTTGGATCAAAGCAACGAAGATTCTGCTTCAA
GTCAACAAGTAACCAAGAAAAGAGCCCAATGGCGGACACTAATTACCATATGAATCCCAA
TCACAACGGTGACCTAGAAGGCGTCACTGGAATGCAAGGATGCCCAAGAGACTAAGAAC
CAGCGACGAGACAATGATGCAGCCAATAAATGCGGATTTTACGCTCCAACGAGAAGCTCAC
GATGAAGATTCTAGAAGAACGGCAAGGGATAAGATCAGATGGTGGCTACCTTTTCATGGG
TAATTTTCGGGCAATACCAATGGATGAGATGTCAAGATTTGATGTAGTCTCAGACCAGGA
GCTCATGGCGCAAAGGTACTCAGGAAACAACAATGGCGTGTCCCTCACGTTAGGTTTACC
TCATTTGTGATAGCTTGTCTCCACGGACCATCAGGGTTTCATGCAGACCCACCATGGGAT
TCCTATAGGGAGAAGAGTGAAAAATAGGAGAAACAGAGGAATATGGACCCGCCACCATCAA
TGGTGGTAGCTCGACCACAACCGCACATTATCAGCGGCAGCTGCCGCGGCTTACAATGG
GATGAACATACAGAACCAGAAGAGATATGTGGCTCAGTTATTGCCCGACTTCGTTGCATA
AACCCTCTCTCTAGAAGGAGAAACCGAAACAGGTTATTATATACGTTTCTAGTTTTTAA
TTAGTATATAGTTTTCTCATACCATTGAACCAAAACAAAGAAATTTAATTTTAGTCT
TTGGTTATATATGGCCGACGGGCTACGTACAGGGCCCTGACGTAGC

>G1589 Amino Acid Sequence (conserved domain in AA coordinates:384-448)

MAAYFHGNPPEISAGSDGGLQTLILMNPTTYVQYTQDNDNSNNNNNSNNNNNTNTNTN
NNNSSFVFLDASHAPQPNASQQFVGIPLSGHEAASITAADNISVLHGYPYPRVQYSLYGS
VDPTHQQAACETPRAQQGLSLTLSSQQQQQQHQQHPIHVFGSGHGEDIIVGSGSTG
SGVTNGIANLVSSSKYLKAAQELLDEVVNADSDDMNAKSQLFSSKKGSCGNDKPVGESSAG
AGGEGSGGGAEEAGKRPVELGTAERQEIQMKAKLNNLHEVEQRYRQYHQQMQMVISSF
EQAAGIGSAKSYTSLALKTISRQFRCLKEAIAQIKANKSLGEEDSVSGVGRFEGSRLK
FVDHHLRQQRALQQLGMIQHPSNNAWRPQRGLPERAVSVLRWLFEHFLHPYPKSDSKHM
LAKQTGLTRSQVSNWFINARVRLWKPMVEEMMEEMKEQAKNMGSMKTPLDQSNEDSAS
KSTSNQEKSPMADTNYHMNPNNHNGDLEGVTMQGPCPKRLRTSDETMMPINADFSSNEKL
TMKILEERQGIKIRSDGGYFPFGNFQYQMDMSRFDVVSQELMAQRYSGNNNGVSLTLGL
PHCDLSLSTDHQGFMTTHGIPIGRRVKIGETEEYGPATINGGSSTTTAHSSAAAAAAYN
GMNIQONQKRYVAQLLPDFVA*

>G160 (38..784)

TCAAATTTGTCATTGTGTTATTCAAATTTTGGAGAAAATGGTGAGAAAGTACCAAAGGTCG
TCAGAAAATAGAGATGAAAAAATGGAAAACGAAAGCAACCTTCAGGTTACTTTCTCAAA
AAGAAGATTCCGGTCTTTTCAAAAAAGCTAGTGAACCTTGCACATTAAGTGGTGCAGAGAT
TCTGTTGATTGTGTTCTCTCCTGGTGGGAAAGTGTTTTCTTTTGGCCATCCAAGTGTTCA
AGAAGTCAATTCATCGCTTTTTCGAATCCTAACCATTAATCTGCCATTGTCCATCATCAGAA
CAACAATCTCCAACCTGTTTGAACCCGTCGCGATAGAAATATCCAATATCTCAACAATAT
ACTCACTGAGGTGCTGGCAAAACAGGAAAAGGAGAAACAGAAGAGAATGGTTTTGGACCT
ATTGAAAGAATCCAGAGAACAGTAGGAACTGGTATGAAAAAGATGTGAAAGATCTCGA
CATGAATGAAACCAACAGCTGATATCTGCTCTTCAAGATGTGAAAAAGAACTGGTAAG
AGAAATGTCTCAATATTCTCAAGTAAATGTTTCGCAGAATTACTTTGGTCAAAGTTCTGG
CGTGATTGGTGGTGGTAAATGTTGGCATTGATCTTTTTGATCAAAGAAGAAATGCATTCAA
CTATAATCCAAACATGGTGTTTCCCAATCATACACCACCAATGTTTGGATACAACATGA
TGGAGTTCTCGTTCGATATCCAACATGAACTACATGTCAAGTTACAACCTCAACCAGAG
CTAGAGTCTGAAGCTAGAAGAACATCCTAATCAATATTTGCGTTATTTTGGCTATGGTTA
CTGTTAGGATTGTTCTTGTATTGTGAGACTTAAGTTTGTTTTTCTTTTAATTTGTTTCA
GTTGGTTGGTTTTTTCATTTTATTTCGTCGTTTGTTCCTTTGTTTGGATATTTTGTGTA
TCCCAGAATAAATTTATTTATCCTTTAAAAA

>G160 Amino Acid Sequence (domain in AA coordinates: 7-62)

MVRSTKGRQKIEMKKMENESNLQVTFKRRFGLFKKASELCTLSGAEILLIVFSPGGKVF
SFGHPSVQELIHRFSNPNNHNSAIVHHQNNNLQLVETRPDRNIQYLNILTEVLANQEKEK
QKRMVLDLLKESREQVGNWYEKDVKDLDMNETNQLISALQDVKKLVREMSQYSQVNVSQ
NYFGQSSGVIGGGNVGIDLDFQRRNAFNYPNMFVFNHTPPMFGYNNDGVLPVISNMNYM
SSYNFNQS*

>G1636 (19..666)

GAGTAATCATCAACGATTATGGCGTCAAGTCAGTGGACGAGGTCGGAGGATAAGATGTTT
GAGCAAGCTTTGGTCTTTTTCTGAAGGATCTCCTAATCGGTGGGAGAGAATCGCTGAT
CAGCTTCATAAATCTGCTGGTGAAGTTAGGGAGCATTACGAGGCTTGGTTCATGATGTT
TTCGAGATTGATTCTGGTCGAGTTGATGTCCCTGATTACATGGATGACTCGGCGGCTGCG
GCGGCGGGTTGGGATTCCGCTGGTCAGATCTCTTTGGGTCTAAACATGGCGAGAGTGAA
CGCAAAAGAGGAACCTCTTGGACAGAGAACGAACACAAATTGTTTCTGATCGGATTAAAG
AGATATGGTAAGGGAGATTGGAGGAGTATCTCGAGAAACGTTGTGGTGACGAGGACACCG
ACGCAAGTCGCGAGTACGCTCAGAAGTATTTTCTGAGACAGAACCTCGGTGAAGAAGGAG
AGGAAAAGGTTCGAGCATCCATGATATAACTACGTTGATGCTACTTTGGCTATGCCTGGG
TCTAACATGGACTGGACTGGCCAAACGCGGAGTCTGTTAGGCGCCGAGCAGCAACAG
ATTATGTCTGAGTTCGGTCAGCAATTGAATCCTGGTCATTTGAGGATTTTGGGTTTCGG
ATGTGATG

>G1636 Amino Acid Sequence (domain in AA coordinates: 100-165)

MASSQWRSEDKMFEQALVLFPEGSPNRWERIADQLHKSAGEVREHYEVLVHDVFEIDSG
RVDVPDYMDDSAAGWDSAGQISFGSKHGSESRKRGTPWTENEHKLFLIGLKRYGKGD
WRSISRNVVVTRPTQVASHAQKYFLRQNSVKKERKRSSIHDITTVDATLAMPGSNMDWT
GQHGSFPVQAPQQQIMSEFGQQLNPGHFEDFGFRM*

>G1642 (1..1077)

ATGGGTCATCACTCATGCTGCAACAAGCAAAAGGTGAAGAGAGGGCTTTGGTCACCTGAA
GAAGACGAAAAGCTCATCAACTACATCAATTCATATGGCCATGGATGTTGGAGCTCTGTT
CCTAAACATGCAGGTTTGCAGAGATGTGGAAAGAGTTGTAGATTAAAGATGGATAAATTAT
CTAAGACCTGATCTTAAACGTGGAAGCTTCTCTCCTCAAGAAGCTGCTCTTATCATTGAG
CTTCACAGCATTTCTGGTAACAGATGGGCTCAAATTGCTAAACATCTACCTGGAAGAACA
GATAACGAGGTCAAGAATTTCTGGAACCTCGAGCATTAAAAAGAAGCTCATGTCTACCAT
CATCACGTCATCATCATCATCTCTCTTCCATGGCGAGTTTGCTCACAAACCTTCTCT
TATCACAATGGATTCAACCTTACTACAGTCGACGATGAAAGTTCAAGATTCATGTCCAAT
ATCATCACAAACACTAACCTTAATTTTCATCACTCCAAGCCATCTCTCTCTCTCTCTCT
CATGTTATGACCCATTGATGTTCCCAACCTCTAGAGAAGGAGATTTCAAGTTTCTAACC
ACAAACAACCCAAACCAATCTCATCACCATGATAATAACCATTACAACAACCTCGACATT
TTGTACCCACACCAACTATAAACAATCATCATCAACCTTCACTTTCTTCTGTCTCAT
GATAATAATCTCCAATGGCCAGCGTTACCAGATTTCCAGCGAGTACCATTTCTGGTTTC
CAAGAAACCTTCAAGATTATGATGATGCTAATAAACTCAACGTGTTTGTGACACCATTTC

AACGATAATGCCAAAAAGTTATTATGTGGAGAAGTTCTCGAAGGCAAAGTACTATCTTCC
 TCCTCACCAATTTCAACAAGATCACGGCCTTTTCTTCCCACCACGTACAACCTTCAAATG
 ACTTCTACGAGTGATCATCAACATCATCATCGAGTGGACTCATACATCAATCACATGATC
 ATACCATCATCATCCTCATCGTCGCCAATCTCTGTGGACAGTACGTCATAACTTAA

>G1642 Amino Acid Sequence (domain in AA coordinates: TBD)

MGHHSCCNKQKVKRGLWSPEEDEKLINYINSYGHGCWSSVPKHAGLQRCGKSCRLRWINY
 LRPDLKRGFSFSPQEAALIELHSILGNRWAQIAKHLPGRTDNEVKNFVNSSIKKLMSSH
 HHGHHHHHLSSMASLLTNLPYHNGFNPTTVDESSRFMSNIITNTPNFITPSHLSLPSP
 HVMTPLMFPTSREGDFKFLTNNPNQSHHHDNNHYNLDILSPTPTINNHHQPSLSLSSCPH
 DNNLQWPALPDFPASTISGFQETLQDYDDANKLNVFVTPFNDNAKLLCCEVLEGGKVLSS
 SSPISQDHGLFLPTTYNFQMTSTSDHQHHRVDSYINHMIIPSSSSSSPISCGQYVIT*

>G1747 (1..777)

ATGAAAATGATGCAAGAGGAGGAAACCGAAAAGGTCCATGGACAGAACAGGAAGACATA
 CTTCTGGTAAATTTTGTTCACCTTATTTGGAGATCGACGATGGGATTTTATAGCAAAAGTA
 TCAGGTTTGAACAGAACAGGAAAGAGTTGCGAGGCTAAGATGGGTTAATTACCTACATCCT
 GGTCTCAAACGTGGCAAGATGACGCCTCAAGAAGAGCGCCTCGTCTTGAGCTTCACGCT
 AAGTGGGGAAACAGGTGGTCGAAAATAGCCCCGAAAATTGCCGGGACGAACGGATAACGAG
 ATAAAGAACTACTGGAGGACTCATATGAGGAAGAAAGCTCAAGAAAAGAAGCGTCTGT
 TCCCCAACCTTCTCATTTTCCAACCTGCAGCTCGTCATCTGTGACCACTACCACCACCAAT
 ACTCAAGATACATCGTGCCACTCGCGTAAATCTTCAGGGGAAGTGAGCTTTACGACACT
 GGAGGTTCCCGATCCACTAGAGAGATGAATCAAGAAAACGAAGACGTGTACTCGTTGGAT
 GATATATGGAGAGAGATTGATCACTCAGCAGTAAACATAATAAAACCGGTTAAAGACATC
 TACTCAGAACAAAGCCATTGCTTAAAGTTACCCAAATCTAGCTTACCATCATGGGAAAGC
 TCATTGGATTCTATATGGAACATGGATGCAGATAAAAGTAAGATATCGTCTTACTTTGCA
 AATGATCAGTTTCTTTCTGTTTCCAACACAGTAGATCACCATGGTCGTCAGGTAA

>G1747 Amino Acid Sequence (domain in AA coordinates: 11-114)

MKMMQEEGNRKGPPWTEQEDILLVNFVHLFGDRRWDFIAKVSGLNRTGKSCRLRWVNYLHP
 GLKRGKMTPOEERLVLELHAKWGNRWSKIARKLPGRTDNEIKNYWRTHMRKKAQEKRPV
 SPTSSFSNCSSSSVTTTTNTQDTSCHSRKSSGEVSFYDTGGSRTREMNQENEDVYSLD
 DIWREIDHSAVNIKPVKDIYSEQSHCLSYPNLASPSWESSLDSIWNMDADKSKISSYFA
 NDQFPFCFQHSRSPWSSG*

>G1749 (59..535)

CAACACTTCTCAGTGACCGTGAGCAACGAATTATTTTTCAGTTCAACGACTCCGCGGAAAT
 GGAAAATTCAGAAAATGTTCCCTCTTACGATCAAAACATCAATTTCACTCCTAATTGAC
 GAGAGATCAAGAACATGTGATCATGGTCTCTGCTTTGCAACAAGTAATATCCAACGTCGG
 AGGTGACACGAACTCGAATGCATGGGAAGCTGATCTTCCACCTTTGAACGCTGGCCCTTG
 TCCTCTTTGTAGTGTACCCGGCTGCTACGGTTGCGTCTTCCCACGACACGAGGCGATAAT
 TAAGAAGGAGAAGAAGCACAAAGGAGTGAGGAAAAAACCATCAGGTAAATGGGCGCGGA
 GATATGGGATCCGAGTTTGAAGTAAGGAGATGGCTTGAACGTTTCCAACAGCGGAGAT
 GCGCGCTAAGGCTTACAACGATGCGGCGCTGAGTTTGTGGAAGAAGATCAGCAAGACG
 TGGCACAAGAAGCGGAGAGGAAGCATCTACCAAGAAGACGACTGAGAAAAATTAACGGAG
 AAGGAGCACGTATAGAAAGGCAGGAAGAGGCATCTTACTTGCTTCACAAGTAAATCAGAA
 TTTTTTTGAAAAGTAAAAACGTTATTTTGTGTTGGTAATAAAATAAAGTAAAAACAAATAT
 TGCTAACGCAAGACTTATCAAGTTCAAGTCGTCGACTGTGAGTGTGTTTTATGTATCTTAC
 TTCATTTTTTGTCTTTCAATTGTGTGTGTGTGT

>G1749 Amino Acid Sequence (conserved domain in AA coordinates: 84-155)

MENSENVPSYDQNIPTFNLTRDQEHVIMVSALQQVISNVGGDTNSNAWEADLPPLNAGP
 CPLCSVTGCGYCVFPRHEALIKKEKHKGVRRKPSGKWAABEIWDPSLKVRRLGTPTAE
 MAAKAYNDAAEFVGRRSARRGTKNGEEASTKKTEKN*

>G1751 (117..923)

AAACACAAACAAAACCTCATATTTTCAATCTCCAGGTGCTTTACACCAACAGAGTCGCAAG
 AAAACAAAACCAAACCTCGGATTTAGTTTGACAGAAGAAGGAATCGAGAGTCGGGTATGC
 ATTATCCTAACAAACAGAACCGAATTCGTCGAGCTCCAGCCCCAACCCGGTATCAAAAGG
 AGCAGTTGTACCCGGAGCAAGAGCTTTCAAGTTATTGTCTCTGCTTTGCAACACGTGATCT
 CAGGGGAAAACGAAACGGCGCGGTGTGAGGTTTTTCAGTGACAGCACAGTGATAAGCG
 CGGGAATGCCTCGGTTGGATTTCAGACACTTGTCAAGTCTGTAGGATCGAAGGATGTCTCG
 GCTGTAACACTTTTTTCGCGCCAAATCAGAGAATTGAAAAGAATCATCAACAAGAAGAAG

AGATTACTAGTAGTAGTAACAGAAGAAGAGAGAGCTCTCCCGTGGCGAAGAAAGCGGAAG
GTGGCGGGAAAAATCAGGAAGAGGAAGAACAAGAAGAATGGTTACAGAGGAGTTAGGCAAA
GACCTTGGGGAAAAATTGACAGCTGAGATCAGAGATCCTAAAAGAGCCACACGTGTTTGGC
TTGGTACTTTTCGAAACCGCCGAAGATGCGGCTCGAGCTTATGATCGAGCCGCGATTGGAT
TCCGTGGGCCAAGGGCTAAACTCAACTTCCCCTTTGTGGATTACACGTCTTCAGTTTCAT
CTCCTGTTGCTGCTGATGATATAGGAGCAAAGGCAAGTGCAAGCGCCAGTGTGAGCGCCA
CAGATTTCAGTTGAAGCAGAGCAATGGAACGGAGGAGGAGGGGATTGCAATATGGAGGAGT
GGATGAATATGATGATGATGATGGATTGTTGGGAATGGAGATTCTTCAGATTTCAGGAAATA
CAATTGCTGATATGTTCCAGTGATAAATGAGCTCTTTCTTGTGGCGTTTTTTGGAGTTA
AGTGCAAGAAGAGATTGACACTGTGGCTTGTTTAAAGTGAACAAGAAAGCAAGCATGT
AATTAGTAGTCTCATCTTTTGTGTTGTGGTCAATTCTATGTTTATCTCATATAAAATCTG
AGTTAAACCTATCTGAGGAGAGAGTAAATAAAGAGGTTAAGAA

>G1751 Amino Acid Sequence (domain in AA coordinates: TBD)
MHYPNNRTEFVGAPAPTRYQKEQLSPEQLSVIVSALQHVISGENETAPCQGFSSDSTVI
SAGMPRLSDSTCQVCRIEGLGCNYFFAPNQRIEKNHQEEIITSSNRRRESSPVAKKA
EGGGKIRKRKNKNGYRGVRQRPWGKFAAEIRDPKRATRVWLGTFFETAEDAAARAYDRAAI
GFRGPRAKLNFPFVDYTSVSSPVAADDIGAKASASVSATDSVEAEQWNGGGGDCNME
EWMNMMMMDFNGDSSDSGNTIADMFG*

>G1752 (25..756)
AAAAAAAAAAAAAAAAAACTTATGGAATATTCCCAATCTTCCATGTATTTCATCTCCA
AGTTCTTGGAGCTCATCACAAGAACTACTCTTATGGAACGAGAGCTGTTTCTTGGATCAA
TCATCTGAACCTCAAGCCTTCTTTGCCCCAATTATGATTACTCCGATGACTTTTCTCA
TTTGAGTCACCGGAGATGATGATTAAAGGAAGAAATTCAAACCGGCGACGTTTCTAACTCC
GAAGAAGAAGAAAAGGTTGGAATTGATGAAGAAAGATCATAAGAGGAGTGAGGAAAAGG
CCGTGGGGGAAATTTGCAGCGGAGATAAGAGATTCAACGAGGAATGGAATTAGGGTTTGG
CTCGGGACATTTGACAAAGCCGAGGAAGCCGCTCTTGCTTATGATCAAGCGGCTTTCGCC
ACAAAAGGATCTCTTGCAACACTTAATTTCCCGGTGGAAGTGGTTAGAGAGTCGCTAAAG
AAAATGGAGAAATGTGAATCTTCATGATGGAGGATCTCCGGTTATGGCCTTGAAGAGAAAA
CATTCTCTTCGAAACCGGCCTAGAGGGAAAAAGCGATCCTCTTCTTCTTCTTCTTCT
TCTAATCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
AAGCAGAGTGTGTGAAGCAAGAAAGTGGTACACTTGTGGTTTTTGAAGATTAGGTGCT
GAGTATTTAGAACAACCTTCTTATGAGCTCATGTTGATCTTGTAATTGATTTCAGCAAAAG
CCACTATTAACCTTTAATTTTGTGATAATTAATCTTGAAATTTGTTTTGTTCTTCTGCA
ATTTCTTTGGTTCTCTTATTTTTTGTGTTGTGATCCAAATGAAATTATGGAAGAGATG
GTGATGTTAAAGTGATATATATATAAAAAAAAAA

>G1752 Amino Acid Sequence (domain in AA coordinates: TBD)
MEYSQSSMYSSPSSWSSSQESLLWNESCFLDQSSEPQAFPCPNYDYSDDFFSFESPEMMI
KEEIQNGDVSNSEEEKVGIDEERSYRGVRKRPWGKFAAEIRDSTRNGIRVWLGTDFKAE
EALALYDQAAFATVDSLATLNFPEVVRESLKKMENVNLHDGGS PVMALKRKHSLRNRPR
GKKRSSSSSSSSSSSSSSSSSTSSSTRSSSKQSVVKQESGTLVVFEDLGAEYLEQLLM
SSC*

>G1763 (33..977)
GTACATTTTTTTTGTATTTTCAGGAACTCCGATGGCGGATCTCTTCCGGTGGTGGCCACG
GCGGCGAGCTTATGGAAGCACTTCAACCTTTTTTACAAAAGTGCTTCCACGTCTGCTTCAA
ATCCTGCGTTTGCCTCTCAAACGATGCGTTTGCCTGCGCCCAAACGACCCATTTTCTT
CTTCTTCTTACTATAATCCTCATGCATCTTCTTCCCTTCACATTCCACAACCACTTACC
CGGATATTTATTTCTGGATCCATGACCTATCCATCTTCATTCCGGGTCGGATCTTCAACAAC
CCGAAAACCTACCAATCTCAGTTCCATTACCAAAACACTATCACTTACACTCACCAAGACA
ACAACACTTGCATGCTCAACTTCATTGAGCCGAGCCAACCGGATTTTATGACCCAACCGG
GTCCGAGTTCGGGTTEGGTTTCAAACCGGCTAAGCTCTATAGAGGAGTGAGGCAAAGAC
ATTGGGGAAAATGGGTGCGCGAGATCCGTTTACCCAGGAACCGAACCCGACTTTGGCTCG
GAACATTCGACACGGCTGAAGAAGCCGCGTTGGCTTATGATCGCGCCGCTTTAAGCTTC
GTGGTGACTCGGCTCGGCTTAACTTCCCAGCTCTCCGATACCAAACCGGCTCGTCTCCGT
CTGACGTTGGCGAATACGGACCTATTCAAGCTGCCGTTGACGCCAAGCTAGAAGCCATAT
TAGCTGAGCCGAAGAATCAGCCGGGCAAAACGGAGAGGACGTCGAGGAAACGAGCTAAAG
CCGCGGCTTCTTTCAGCTGAGCAGCCGTCAGCGCCACAACAACATTCGGGTCGGGTGAAA
GTGATGGGTGCGGTTTACCGACTTCGGATGTTATGGTGCAGGAGATGTGCCAAGAGCCAG

AGATGCCATGGAATGAAAATTTTCATGCTCGGCAAGTGTCTTCTTATGAGATAGATTGGG
CTTCAATTTTATCGTGAAAAATTAGGATTCATTTTATTCATTTTAACTTGTTG
TATTTTCTTTAACTTTAGGGTTATTAGCTGTGCGTAAATTTGTAATTTAGCATTTTG
TATGAATGTAATGCAAGTGTGTAAATTATGGACAGCTCAAGCTTTTTTGTAAAA

>G1763 Amino Acid Sequence (conserved domain in AA coordinates:140-209)

MADLFGGGHGGELMEALQPFYKSASTSASNPFASSNDAFASAPNDPFSSSSYNPHASF
FPSHSTTTYPDYSGSMTYPSFSGDLQQPENYQSQFHYQNTITYTHQDNNTCMLNFI
SQPDFMTQPGPSSGSVSKPAKLYRGVQRHWGKWVAEIRLPRNRLWLGTFTDTEEAAL
AYDRAAFKLGRDSARLNFPALRYQTGSSPSDVGEYGPIQAAVDKLEAILAEPKNQPGKT
ERTSRKRAKAAASSAEQPSAPQQHSGSGESDGSPTSDDVMVQEMCQEPMPWNENFMLG
KCPSYEIDWASILS*

>G1766 (32..1216)

AGGCTATTCTCGAAAAACAAGAATAAAGAATGAATTCGTTTTCACAAGTACCTCCTGG
CTTCAGATTTTCATCCTACTGATGAAGAACTTGTAGACTACTACTTGAGGAAAAAGTTGC
ATCAAAGAGAATAGAAATCGATATCATCAAGGATGTTGATCTTTACAAGATTGAGCCATG
TGATCTTCAAGAGTTATGCAAGATAGGAAACGAAGAGCAGAGCGAATGGTACTTCTTAG
TCATAAAGACAAGAGTATCCACGGGAACTCGAACCAATAGAGCCACGAAAGCAGGATT
TTGGAAAGCCACTGGAAGAGACAAGGCTATATATATAAGACATAGTCTTATCGGTATGAG
GAAACACTTGTGTTTTACAAAGGAAGAGCCCCAAATGGTCAGAAATCCGATTGGATCAT
GCACGAATATCGCTTAGAAACAAGTGAATGGAACCCCTCAGGAAGAAGGATGGGTAGT
ATGTAGGGTATTCAAGAAGAAATTTGGCAGCGACAGTGAGGAAATGGGAGATTACCATTC
ATCACCATCGCAGCATTTGGTACGATGATCAGCTCTCTTTTATGGCCTCCGAGATCATTC
TAGTTCTCCACGACAGTTTCTTCCAATCATCATTATAACCGCCACCATCACCAGCAGAC
ATTGCCTTGTGGCCTCAATGCATTCAACAACAACATCTAAGTCAATGCAAGCAAGA
GCTCGAGTTACATTACAATCAATGGTACAACATCAACAACAAACCATCATCTTCGTGA
ATCTATGTTTCTCCAGCTTCCTCAGCTCGAAAGCCCTACCAGTAATTGCAATTCTGACAA
CAACAATAACACAAGAAATATTAGTAAC'TGCAGAAATCATCAAATATATCTCATGAGGA
ACAATTGCAACAAGGAATCAAAGTTTACGCTCTCTGTATTACGATCAAGGAGTAGAGCA
AATGACTACTGACTGGAGAGTTCTCGATAAATTTGTTGCTTCACAGCTTAGCAATGATGA
AGAGGCTGCAGCCGTGGTTTCTTCTTCTTCTCATCAAAACAACGTCAAGATTGACACGAG
AAACACGGGTTATCATGTGATAGATGAGGGAATAAATTTGCCGAGAAATGATTCTGAAAG
GGTTGTTGAAATGGGAGAAGAGTATTCAAATGCTCATGCTGCTTCTACTTCTTCAAGTTG
TCAGATTGATCTCTAGAAATAGTGATAGAGAGATGAAAAAGATGCAAGGTGAATATATAT
GAAATACATGCACACTAGTGTTATTTATACCTTAAAGATGGAAGGGGAAAAACAAGGAGT
TATTTCTCGGATTTATGGAGGTTTTGTACATAATAAAACCTACAACCATATGGTATTTT
CTTTTGAAAAA

>G1766 Amino Acid Sequence (domain in AA coordinates: 10-153)

MNSFSQVPPGFRFHPTEDELVDYLRKKVASKRIEIDI IKDVDLYKIEPCDLQELCKIGN
EEQSEWYFFSHDKKYP TGTTRTNRA TKAGFWKATGRDKAIYIRHSLIGMRKTLVFKGRA
PNGQKSDWIMHEYRLTSENGTPQEEGWVVC RVFKKLAATVRKMGDYHSSPSQHWYDDQ
LSFMASEI IISSPRQFLPNHHYNRHHHQQTLP CGLN AFNNNNPNLQCKQELHLHYNQMVQ
HQQNHHLRESMFLQLPQLESPTSNCNSDNNNNTRNINSLQKSSNISHEEQLQQGNQSFS
SLYDQGV EQMTTDWRVLDKFV ASQLSNDEEAAV VSSSSHQNNVKIDTRNTGYHVIDEG
INLPENDSERVEMGEEYSNAHAASTSSSCQIDL*

>G1767 (1..1596)

ATGGATACTCTCTTTAGACTAGTCAGTCTCCAACAACAACAATCCGATAGTATCATT
ACAAATCAATCTTCGTTAAGCAGAACTTCCACCACCACTACTGGCTCTCCACAACTGCT
TATCACTACAACCTTCCACAAAACGACGTCGTCGAAGAATGCTTCACTTTTTTCATGGAT
GAAGAAGACCTTCTCTTCTTCTTCTCACCACAACCATCACAACCACAACAATCCTAAT
ACTTACTACTCTCTTTCACTACTCCCAACCAATACCATCCCGCCACATCATCAACCCCT
TCCTCCACCGCCGAGCCGAGCTTTAGCCTCGCCTTACTCCTCCTCCGCGCCACCATAAT
GACCTTCCGCGTTCTCCATACCTCAAACCTCCTCGTCTCTCGACTTCTCAGCCAATGCC
AAGTGGGCAGACTCGGTCTTCTTGAAGCGGCAGCTGCCTTCTCCGACAAAGACACTGCA
CGTGCGCAACAAATCCTATGGACGCTCAACGAGCTCTTCTCCGTACGGAGACACCGAG
CAAAACTGGCTTCTTACTTCTTCAAGCTCTTCAACCGCATGACCGGTTCAAGCGAA
CGATGCTACCGAACCATTGGTAACAGCTGCAGCCACAGAGAAGACTTGCTCTTTCGAGTCA
ACGCGAAAACTGTACTAAAGTTCCAAGAAGTTAGCCCTGGGCCACGTTTGGACACGTG

GCGGCAAACGGAGCAATCTTGAAGCAGTAGACGGAGAGGCAAAGATCCACATCGTTGAC
ATAAGCTCCACGTTTTTGCACTCAATGGCCGACTCTTCTAGAAGCTTTAGCCACAAGATCA
GACGACACGCCTCACCTAAGGCTAACACAGTTGTCTGGCCAAACAAGTTTGTCAACGAT
CAAACGCGCTCGCATCGGATGATGAAAGAGATCGGAAACCGAATGGAGAAATTCGCTAGG
CTTATGGGAGTTCTTTCAAATTTAACATTATTCATCACGTTGGAGATTATCTGAGTTT
GATCTCAACGAACCTCGACGTTAAACCAGACGAAGTCTTGGCCATTAAGTGCCTAGGCGCG
ATGCATGGGATCGCTTACGTTGGAAGCCCTAGAGACGCTGTGATATCGAGTTTCCGACGG
TTAAGACCGAGGATTGTGACGGTCGTAGAAGAAGAAGCTGATCTTGTGCGAGAAGAAGAA
GGTGGCTTTGATGATGAGTTCTTGAGAGGGTTTGGAGAATGTTTACGATGGTTTAGGGTT
TGCTTCGAGTCATGGGAAGAGAGTTTTCAGGACGAGCAACGAGAGGTTGATGCTAGAG
CGTGCACGGGGACGTGCGATCGTTGATCTTGTGGCTTGTGAGCCGTCGGATTCCACGGAG
AGGCGAGAGACAGCGAGGAAGTGGTCGAGGAGGATGAGGAATAGTGGGTTTGGAGCGGTG
GGGTATAGTGATGAGGTGGCGGATGATGTCAGAGCTTTGTTGAGGAGATATAAAGAAGGT
GTTTGGTCGATGGTACAGTGTCTGATGCCGCCGAATATTCCTTTGTTGGAGAGATCAG
CCGGTGGTTTGGGCTAGTGGCTGGCGGCCAACGTAA

>G1767 Amino Acid Sequence (domain in AA coordinates: 255-272)

MDTLFRLVSLQQQQSDSIITNQSSLRTSTTTGSPQTAYHYNFPQNDVVEECFNFFMD
EEDLSSSSSHNHNHNPNNTYSPFTTPTQYHPATSSTPSSTAAAAALASPYSSSGHHN
DPSAFSIPQTPPSFDFSANAKWADSVLLEAARAFSDKDTARAQQILWTLNELSSPYGDTE
QKLASYFLQALFNRMGTSGSERCYRMTVTAATEKTCSESTRKTVLKFQEVSPWATFGHV
AANGAILEAVDGEAKIHIVDISSTFCTQWPTLLEALATRSDDTPhRLRTTVVVKFVND
QTASHRMKEIGNRMEKFARLMGVPFKFNIHHVGDLSFDLNELDVKPDEVLAINCVA
MHGIASRGSPRAVISSFRRLRPRIIVTVVEEADLVGEEGGFDDEFRLRGFGECLRWFVRV
CFESWEESFPRTSNERLMLERAAGRAIVDLVACEPSDSTERRETARKWSRRMRNSGFGAV
GYSDEVADDVRALLRRYKEGVMSMVQCPDAAGIFLCWRDQPVVWASAWRPT*

>G1778 (1..627)

ATGATGGGATACCAAACTCTAATTTCTCCATGTTTTTTTCTCGGAAAATGACGAC
CAAAACCACCACAACCTACGATCCTTATAATAATTTCTCTTCATCAACTTCTGTTGATTGC
ACTCTCTCACTTGGAAACCCCTACTCGTCTCGACGACCACCATAGATTTTCTTCTGCT
AATTCTAACAACATCTCCGGCGACTTTTATATTACGGAGGAAACGCTAAGACTTCTTCG
TACAAGAAGGGTGGTGTGCTCATAGCCTACCTCGCCGTTGTGCTAGCTGCGACACCACT
TCAACTCCTCTATGGAGAAACGGACCAAAAGGACCTAAGTCGTTATGTAACGCGTGTGGA
ATCCGATTCAAGAAAGAGGAGAGGCGTGGCAGCGCCAGAACTTAACGATCTCCGGTGGG
GGTTCATCAGCGGCAGAGTCCAGTAGAGAATTCGTACAACGGAGGTGGAACTATTAC
AGTCATCATCATCACTATGCCTCGTCTCGCCGTCGTGGCTCATCAGAACACACAA
AGAGTTCCATATTTCTACCGGTTCCGGAGATGGAATATCCCTACGTGGATAACGTCACG
GCTTCTTCTTTTATGCTTTGGAATTGA

>G1778 Amino Acid Sequence (domain in AA coordinates: 94-119)

MMGYQTNSNFSMFFSENDDQNHNYDPYNNFSSSTSVDCITLSLSTPSTRLDHHRFSSA
NSNNISGDFYIHGNAKTSSYKGGVAHSLPRRCASCDTTSTPLWRNGPKGPKSLCNACG
IRFKKEERRATARNLTISGGGSSAAEVPVENSYNGGNYSHHHHHYASSSPSWAHQNTQ
RVPYFSPVPEMEYPYVDNVTASSFMSWN*

>G1789 (108..413)

CAAGGACTCTGCGACATCTGTGCAACATATCATTTCTCTCAGAATCTCTTTCTTTCTAGG
TTTATTACTACACAAACCAACATCATCAACTTTAGTTACTAAACAATGGCATCAGGCT
CAATGTCTTCTTATGGCTCTGGCTCATGGACTGTTAAGCAGAACAAAGCCTTTGAGCGTG
CTCTAGCAGTCTATGACCAAGACACTCCGGACCGTTGGCACAATGTTGCTAGAGCTGTTG
GTGGTAAACACCAGAAGAAGCTAAGAGACAGTATGACCTTCTAGTTCTGTGACATCGAAA
GCATCGAGAATGGTCACGTGCCATTCCCTGACTACAAGACTACTACAGGAAACAGCAACA
GAGGCAGGCTGCGTGATGAGGAAAAGAGGATGAGAAGCATGAAGCTGCAGTGAGACAAGA
AGCAACAAAACCTAACTACGTATGATCGTCAAAATAAAAGAGAATCACTTCAGAGAGATG
TGTTTTTTTCAATGTCTGACGAATCAATGTTTTTTTCTTGCAATTTCTCATGTTTTTCCC
TAAGAAATGGTTTTTTTTCGAGGCAACAAAAA

>G1789 Amino Acid Sequence (domain in AA coordinates: 1-50)

MASGSMSSYSGSWTVKQNKAFERALAVYDQDTPDRWHNVARAVGGKTPPEAKRQYDLLV
RDIESIENGHVFPDYKTTTGNNSNRGLRDEEKRMRSMKLQ*

>G1790 (63..1346)

GAAAAAGACTTCACTTTTTTTTTTACTAATTAATTAGTTTTTTTTTCTCCTTTCCAAAA
 CAATGGAGAATTTTCGTCGACGAGAATGGTTTTGCTTCTCTAAACCAAAACATCTTCACAC
 GTGATCAAGAACACATGAAAGAAGAAGATTTTCCATTTCGAAGTCGTCGACCAATCAAAAC
 CTACAAGCTTTCTTCAAGATTTTCACCATCTTGATCATGATCATCAGTTTGATCATCATC
 ATCATCATGGCTCCTCATCTTCACATCCTTTGCTCAGCGTCCAAACTACGTCTTCTTGTA
 TCAATAATGCTCCTTTTCGAGCATTGCTCTTACCAAGAAAACATGGTCGATTTCTATGAAA
 CTA AACCAAAATTTGATGAATCATCATCATTTTCCAAGCAGTGGAAAACATCACTTCACTC
 GTAATCATCATCATCATCAAGAGATCAATTTGGTCGATGAACATGATGATCCTATGGACT
 TGGAGCAAAACAACATGATGATGATGAGGATGATCCCTTTTGATTACCTCTCTACAGAGA
 CTTTCAAACCTATGAACTTCGTAATGCCAGATGAAATTTTCATGTGTTTCTGCAGATAATG
 ATTTGTTATAGAGCAACGAGTTTCAACAAGACCAACCATTTCTTACACGAAAGTTGTCTT
 CTTCTTCTTCATCATCATCATGGAAGAAAACCAAAAAGTCAACCTTAGTCAAAGGACAAT
 GGACTGCTGAAGAAGACAGGGTACTGATTCAACTCGTGGAGAAGTATGGATTGCGTAAAT
 GGTCGCATATCGCTCAAGTGTACCGGGAAGAATCGGGAACAATGTAGAGAGAGGTGGC
 ATAACCATTTGAGACCTGACATTAAAGAAAAGAACATGGAGTGAAGAAGAGGACAGAGTGT
 TGATAGAATTTCACAAAGAGATTGGAACAAATGGGCAGAGATTGCGAAAAGACTCCCGG
 GAAGAACAGAGAAGCTCGATCAAGAACCATTGGAACGCAACAAAAGAAGACAATTCTCTA
 AAAGAAAATGTAGATCTAAGTATCCAAGACCTTCTCTGTTGCAGGATTACATCAAGAGCT
 TGAATATGGGAGCTTTGATGGCTTCTTCTGTTCTGCAAGAGGTAGACGCAGAGAGAGTA
 ATAACAAGAAGAAGGATGTTGTTGTTGCGGTTGAGGAGAAGAAGAAGGAAGAGGAGGTGT
 ATGGACAAGACAGGATTGTGCCTGAATGTGTGTTTACTGATGATTTTGGATTCAATGAGA
 AGCTGCTTGAGGAAGGATGTAGCATTGACTCTTTGCTTGATGACATTCCTCAGCCTGACA
 TTGATGCTTTTGTGTTTATGGCTCTGATTTGTATTTTATTCTGCTTGTTCAGTTTTGT
 TGTTTTTTGTGTTGCTTTTTTATACGAGACAGATTCCACCAAACTTCAATAATTTGAAAAG
 ATATAAAATATTTTGCTTTTTTAAAAAAAAAAAAAAAAAAAAA

>G1790 Amino Acid Sequence (conserved domain in AA coordinates:217-316)

MENFVDENGFASLNQNIIFTRDQEHMKEEDFPFEVVDQSKPTSFLQDFHHLDDHDHDFDHHH
 HHGSSSSHPLLSVQTTSSCINNAPFEHCSYQENMVDYETKPNLMNHHHFQAVENSFYFTR
 NHHHHQEINLVDEHDDPMDLEQNMMMMRMIPFDYPPTETFKPMNFVMPDEISCVSADND
 CYRATSFNKTKPFLLTRKLSSSSSSSSWKETKKSTLVKGQWTAEDRVLIQLVEKYGLRW
 SHIAQVLPGRIGKQCRERWHNHLRPDIKKETWSEEDRVLIEFHKEIGNKWAEIAKRLPG
 RTENSINKHWNATKRRQFSKRKCRSKYPRPSLLQDYIKSLNMGALMASSVPARGRRRESN
 NKKKDVVVAVEEKKKEEVYQDRIVPECVFTDDPGFNEKLLLEEGCSIDSLLDDIPQPD
 I DAFVHGL*

>G1791 (36..455)

ATGTACATGCAAAAACAAAACCTTAAAGCTTTTCATGGAACGTATAGAGTCTTATAACA
 CGAATGAGATGAAATACAGAGGCGTACGAAAGCGTCCATGGGGAAAATATGCGGCGGAGA
 TTCGCGACTCAGCTAGACACGGTGCTCGTGTGTTGGCTTTGGGACGTTTAAACACAGCGGAAG
 ACGCGGCTCGGGCTTATGATAGAGCAGCTTTCGGCATGAGAGGCCAAAGGGCCATTCTCA
 ATTTTCTCTCAGGATATCAAATGATGAAGGACGGTCCAAATGGCAGCCACGAGAATGCAG
 TGGCTTCTCTCGTCTCGGGATATAGAGGAGGAGGTGGTGGTGATGATGGGAGGGAAGTTA
 TTGAGTTTCGAGTATTTGGATGATAGTTTATTGGAGGAGCTTTTAGATTATGGTGAGAGAT
 CTAACCAAGACAATTGTAACGACGCAACCGCTAGATCATCACTACTTACTTACAGTGTA
 ATGTTTTTGGAGTAAAGAGTAATAATCAATATAATACTTTAGTTTAGGAAAAA
 AAAAAA

>G1791 Amino Acid Sequence (domain in AA coordinates: TBD)

MERIESYNTNEMKYRGVRKRPWGKYAAEIRDSARHGARVWLGTFTNTAEDAARAYDRAAFG
 MRGQRAILNFPHEYQMMKDGPNNGSHENAVASSSSGYRGGGGDDGREVIEFEYLLDDSLLE
 ELLDYGERSNQDNCNDANR*

>G1793 (59..1783)

AGTGATTTATTGATTAAACCAACACAAAATAAACAGATTTGACTCAAAAAGAAGAAAAT
 GAATTCCTAACAACCTGGCTTGGCTTTCCTCTTTCACCGAACAACCTCTTCTTGCCTCCTCA
 TGAATACAACCTTGGCTTGGTCAGCGACCATATGACAACCTTTTCAAACACAAGAGTG
 GAATATGATCAATCCACACGGTGGAGGAGGAGATGAAGGAGGAGAGGTTCCAAAAGTGGC
 CGATTTTCTCGGTGTGAGCAACCGGACGAAAACCAATCCAACACCTAGTAGCTTACAA
 CGACTCAGACTACTACTTCCATACCAATAGCTTGATGCCTAGCGTCCAATCAAACGATGT
 CGTTGTAGCAGCTTGTGACTCCAATACTCCTAACACAGTAGCTATCATGAGCTTCAAGA

GAGTGCTCACAATCTACAGTCACTTACTTTGTCCATGGGGACCACCGCTGGTAATAATGT
TG TAGACAAAGCTTCACCATCCGAGACCACCGGGGATAACGCTAGCGGTGGAGCACTAGC
CGTTGTTGAGACGGCCACGCCAAGACGTGCATTGGACACTTTCGGACAACGAACCTCGAT
CTATCGTGGTGTCAACAAGACATCGATGGACTGGTCGATATGAGGCTCATCTATGGGATAA
TAGTTGTAGAAGGGAAGGCCAGTCTAGGAAAGGAAGACAAGTTTACTTGGGTGGATATGA
CAAAGAAGATAAAGCAGCAAGATCATATGATCTAGCTGCACTTAAGTACTGGGGTCCCTTC
AACTACTACTAATTTCCCATACAACTACGAGAAAGAGTAGAGGAAATGAAGCACAT
GACGAGACAAGAGTTCGTGGCTGCCATTAGAAGGAAAAGTAGTGGATTTTCGAGAGGCGC
TTCGATGTATCGAGGAGTTACAAGGCATCACCAACATGGAAGATGGCAAGCAAGGATCGG
CCGAGTCGCCGGAACAAAGACCTCTACTTGGGAACCTTTAGCACTGAGGAAGAAGCAGC
AGAAGCTTACGATATAGCTGCAATAAAGTTTAGAGGACTTAATGCAGTGACCAACTTCGA
GATCAACCGGTACGACGTGAAAGCCATTCTAGAGAGTAGCACTCTTCCCATCGGAGGAGG
CGCAGCTAAACGGCTCAAAGAAGCTCAAGCTCTTGAGTCTTCAAGGAAACGCGAGGCGGA
GATGATAGCCCTTGGTTCAAGTTTCCAGTACGGTGGTGGCTCGAGCACAGGCTCTGGCTC
CACCTCATCAAGACTTCAGCTTCAACCTTACCTCTAAGCATTCAACAACCATTAGAGCC
TTTTCTATCTCTTCAGAACAAATGACATCTCTCATTACAACAACAACATGCTCACGATTC
CTCCTCTTTTAATCACCATAGCTATATCCAGACACAACCTTCATCTCCACCAACAGACCAA
CAATTACTTGCAGCAACAGTCGAGCCAGAACCTCAGCAGCTCTACAATGCGTATCTTCA
TAGCAATCCGGCTCTGCTTCATGGACTTGTCTCTACCTCTATCGTTGACAACAATAATAA
CAATGGAGGCTCTAGTGGGAGCTACAACACTGCAGCATTTCTTGGGAACACGGTATTGG
TATTGGGTCCAGCTCGACTGTTGGATCGACCGAGGAGTTTCCAACCGTTAAAACAGATTA
CGATATGCCCTTCAGTGATGGAACCGGAGGGTATAGTGGTTGGACCAGTGAGTCTGTTCA
GGGGTCAAACCTGGTGGTGTTTTCACTATGTGGAATGAGTAAACAAGGATCTCTTTCTT
GCGGCACAAGGAATGGGT

>G1793 Amino Acid Sequence (conserved domain in AA coordinates:179-255, 281-349)

MNSNNWLGFPLSPNNSSLPPHEYNLGLVSDHMDNPFQTQEWNMINPHGGGDEGGEVVKV
ADFLGVSKPDENQSNHLVAYNDSYFHTNSLMPVSQSNVVAACDSNTPNNSSYHELQ
ESAHLQSLTSLMGTAGNNVVDKASPSETTGDNASGGALAVVETATPRRALDTFGQRTS
IYRGVTRHRWTRGYEAHLWDNSCRREGQSRKGRQVYLGYYDKEDKAARSYDLAALKYWGP
STTTNFPITNYEKEVEEMKHMTRQEFVAAIRRKSSGFSRGASMYRGVTRHHQHGRWQARI
GRVAGNKDLYLGTFFSTEEEAEEAYDIAAIKFRGLNAVTFEINRYDVKAILESSTLPIGG
GAAKRLKEAQALESRRKREAEMIALGSSFPQYGGGSTSGSGSTSSRLQLQPYPLSIQPPLE
PFLSLQNNDISHYNNNNNAHDSSSFNHHSYIQTQLHLHQQTNNYLQQQSSQNSQQLYNAYL
HSNPALLHGLVSTSIVDNNNNGGSSGSYNTAAFLGNHIGIGSSSTVGSTEEFPTVKTD
YDMPSSDGTGGYSGWTSSESVQGSNPGGVFTMWNE*

>G1795 (27..422)

ACAAACACGCAAAAAGTCATTAATATATGGATCAAGGAGGTGCGAGGTGTCGGTGCCGAGC
ATGGAAAGTACCGGGGAGTTCCGAGACGACCTTGGGGAAAATATGCAGCAGAGATACGAG
ATTTCGAGGAAGCACGGTGAACGTGTGTGGCTTGGAAACGTTCCGATACGGCAGAGGAAGCGG
CTAGAGCCTATGACCAAGCTGCTTACTCCATGAGAGGCCAAGCAGCAATCCTTAACTTCC
CTCATGAGTATAACATGGGGAGTGGTGTCTCTTCTTCCACCGCCATGGCTGGATCTTCTCCT
CCGCTCCGCTCCGCTTCTTCTTCTTCTTAGGCAAGTTTTTGAATTTGAGTACTTGGATG
ATAGTGTTTTGGAGGAGCTCCTTGAGGAAGGAGAGAAACCTAACAAGGGCAAGAAGAAAT
GAGCGAGATATAATTCATGATTATTTCTAA

>G1795 Amino Acid Sequence (domain in AA coordinates: 12-80)

MDQGGRGVGAEHGKYRGVRRRPWGKYAAEIRDSRKHGERVWLGTFTDAEEAARAYDQAAAY
SMRGQAAILNFPHEYNMGSVSSSTAMAGSSSASASASSSRQVFEEFYLDSDVLEELLE
EGEKPNGKGGK* —

>G1800 (61..894)

CCATTATCATATCCTCTTCTTCTTCTTCACTATCAATCTTCTTCTTCCACTACAACACAA
ATGGAGAAATCATCCTCAATGAAACAATGGAAGAAGGGTCTGCTCGGGGTAAAGCGGT
CCACAAAACGCTCTTTGTCACTACCGTGGAGTCAGGCAAAGGACTTGGGGCAAATGGGTG
GCTGAGATCAGAGAGCCCAAGAAGAGGGCAAGACTTTGGCTTGGCTCTTTCGCTACAGCT
GAAGAAGCAGCTATGGCTTATGATGAGGCTGCCTTGAAACTCTATGGGCACGACGCATAC
CTCAACTTACCTCATCTTCAGCGGAATACAAGACCTTCTCTGAGTAACTCTCAGAGGTTT
AAATGGGTACCTTCAAGGAAGTTTATATCTATGTTTCTTCTCATGTGGTATGCTAAACGTG
AATGCTCAGCCTAGTGTTCACATAATCCAGCAAAGACTAGAAGAACTCAAGAAAACCTGGA

CTTTTATCTCAATCCTATTCTTCTAGTTCTTCCTCCACCGAATCAAAAACTAATACTAGC
 TTTCTTGATGAGAAGACCAGCAAGGGAGAAACAGACAATATGTTTCGAAGGTGGTGATCAG
 AAGAAACCAGAGATCGACCTGACCGAGTTTCTTCAGCAACTAGGAATCTTGAAGGATGAA
 AATGAAGCAGAACCAAGTGAGGTAGCAGAGTGTTCATTCCTCCACCATGGAACGAGCAA
 GAAGAACTGGAAGTCCTTTTCAGAACTGAGAATTCAGCTGGGATACCTGATCGAGATG
 CCAAGAAGTGAAACCACAACATATGCAATTTGACTCCAGCAACTTCGGAAGCTATGATTTT
 GAGGATGATGTATCCTTCCCTTCCATCTGGGACTACTACGGAAGCTTAGATTGAGTAAAA
 GCAATTTAAGGTAGATCAAGATTGAGAAGTACACAAATGGTTTTGGATTAGTGTAGCGT
 TTTGGAAAAGAGACATAGGTAGTGAGAGTGCAGTCTTTTATTATGCAGCAATAAAGTGAG
 TCAGTGATACACCGAGTTGTTCGCTTTTTTTGGTATATTAATGAAGCATGTTCATTTTTT
 CGCTAAAAAAAAAAAAAAAAAAAAA

>G1800 Amino Acid Sequence (domain in AA coordinates: TBD)
 MEKSSSMKQWKKGPARGGKGPQNALCQYRGVVRQRTWGWKVAEIREPKRRLWLWLSFATA
 EEAAMAYDEAALKLYGHDAYLNLPHLQRNTRPSLSNSQRFKWPSPRKFSMFPSGMLNV
 NAQPSVHIIQQRLEELKKTGLLSQSYSSSSSSTESKTNTSFLDEKTSKGETDNMFEGGDQ
 KKPEIDLTEFLQQLGILKDENEAEPSVAECHSPPPWNEQEETGSPFRTENFSDTLIEM
 PRSETTTMQFDSSNFGSYDFEDDVSPFPIWDYYGSLD*

>G1806 (1..1356)
 ATGCAGAGCAGCTTCAAACCGTTCTCTTCACTCTGATTTCTACTCTCAATCCTCTTAC
 TTCTTCAGAGGAGATAGTTGTCTTGAGGAGTTTCATCAACCAGTCAATGGTTTTCCACCAT
 GAAGAAGCTATCGATTAAAGTCCAAATGTCACTATTGCTTCAGCTAATTACACTACACG
 ACGTTTGATACGGTTATGGATTGTGGTGGTGGTGGTGGTGGCTTGAGGGAGAGACTTGAA
 GGAGGAGAAGAGGAGTGTGTTGGACACAGGGCAATTAGTGTACCAGAAAGGGACAAGATTA
 GTAGGAGGAGGAGTAGGAGAAGTGAACAGCAGTGGTGTGATTCGGTTTCAGCTATGGCT
 GATAACAGTCAACATACTGACACTTCCACAGATATTGATACTGATGACAAGACTCAGTTG
 AATGGAGGTCAACAGGGATGCTATTGGCTACAAATTGTTTCAGATCAATCCAATGTGAAA
 TCTAGTGATCAAAGGACACTTCGTCGACTTGCTCAGAACCGGGAGGCTGCTAGGAAAAGT
 CGGTTGAGGAAAAAGGCCTATGTTTCAGCAACTTGAGAATAGTCAATCAGGCTTGACAG
 CTAGAGGAAGAGCTCAAAGAGCTCGCCAACAGGGATCTTTGGTTGAAAGAGGAGTTTCA
 GCGGATCACACGCATTGGCAGCAGGAAATGGTGTCTTTTCATTGGAATTGGAATATACA
 CGTTGGAAGGAGGAACATCAAAGAATGATCAACGACTTAAGATCGGGTGTGAATTTCGAG
 TTAGGTGACAACACTACGCGTTCTAGTGGATGCTGTGATGAGTCACTATGATGAAATA
 TTCAGGCTAAAGGGAATTGCACTAAAGTTGAAGTCTTTCATATGCTCTCAGGCATGTGG
 AAGACACCTGCCGAGAGATTTTTCATGTGGTTAGGTGGATTTAGATCATCAGAGTTACTT
 AAGATATTGGGGAACCATGTGGATCCATTGACGGACCAGCAGTTGATAGGCATTTGCAAC
 CTTTCAGCAATCGTCTCAACAAGCAGAGGATGCATTGTACAAGGCATGGAAGCTCTACAA
 CAATCACTTCTCGAGACGCTTCTCTGCTTCTATGGGTCCAACTCTTCAGCAAATGTT
 GCAGATTATATGGGTATATGGCTATGGCTATGGGCAAACCTTGGCACTCTTGAAAACCTC
 TTTCGCCAGGCTGATTTATTGAGGCAACAACTCTGCAACAGCTTCACAGAATTCTCACC
 ACACGACAAGCTGCTCGCGCTTTTGGTTCATCCACGATTATATTTCTCGGCTTAGAGCA
 CTTAGCTCTCTATGGTTAGCCAGACCTAGAGACTAA

>G1806 Amino Acid Sequence (domain in AA coordinates 165-225)
 MQSSFKTVFPFPDFYSQSSYFFRGDSCLEEFHQPVNGFHHHEAIDLSPNVTIASANLHYT
 TFDTVMDCGGGGGGLRERLEGGEEELDTGQLVYQKGTSLVGGGVGEVNSSWCDSDVSAMA
 DNSQHTDTSTDIDTDDKTQLNGGHQGMILLATNCSQSNVKSSDQRTLRLRLAQNREAAARKS
 RLRKKAYVQQLENSRIRLAQLEELKRARQQGSLVERGVSDHHTHLAAGNGVFSFELEYT
 RWKEEHQRMINDLRSGVNSQLGDNLDRLVLVDVMSHYDEIFRLKGIGTKVEVFHMLSGMW
 KTPAERFFMWLGGFRSSSELLKILGNHVDPLTDQQLIGICNLQSSQQAEDALSQMEALQ
 QSLLETLSASMGPNSSANVADYMGHMAMMGLGTLENFLRQADLLRQQLQQLHRIIT
 TRQARAFLVIHDYISRLRLSSLWLARPRD*

>G1811 (93..827)
 AAAGGAGCATTGGTATCTCAACAATATTTGCCCTTTCTCTATCTCTCTCATCACTAT
 TTGCCATCTCTTTCTCTCTCCCTCTCTTTCAAATGTCAATAAACCAATACTCAAGCGATT
 TCCACTACCATTCTCTCATGTGGCAACAACAGCAGCAACAACAACACCAAAAACGACG
 TCGTGAAGAAAAAGAAGCTCTTTTCGAGAAACCCCTTAACCCCAAGTGACGTCGGAAAAC
 TCAACCGCCTCGTCATCCAAAACAGCACGCCGAGAGATACTTCCCACTAGCGGCCGCCG
 CCGCAGACGCCGTGGAGAAAGGACTTCTCCTCTGCTTTGAGGACGAGGAAGGTAAACCAT

GGAGATTCAGATACTCGTACTGGAACAGTAGCCAGAGTTATGTCTTGACCAAAGGCTGGA
GCAGATACGTCAAGGAGAAGCACCTTGACGCCGAGAGCTCGTTCTCTCCATCGACACC
GTTGAGACGGCGGAAGATTCTTCATTGGCTGGAGAAGACGCGGTGACTCTTCTCTCCTCCT
CCGACTCTTATCGCCATGTTCAATCCAATGCCTCGCTCCAATATTATCCTCATGCAGGGG
CTCAAGCGGTGGAGAGCCAAAGAGGCAACTCGAAGACATTAAGACTGTTCCGAGTGAACA
TGGAGTGCCAGCTAGATTCCGACTGGTCCGAGCCATCCACACCTGACGGTTCACACAT
ATACAACCAATCAGCAGCAGTTTCTATTCTACCTCAACAACAACACTATCCTCCTCCGT
ACTACATGGACATAAGTTTACAGGAGATATGAACCGGACGAGCTAGAAGCCCACAAGGA
TAAAAAAAAGCTTCACATCTGGTCTGTATGTTGTCATAGATGTTGATTCTCTTAATTT
TACACAAGCTTCATTTTGCATTATTAAAGTAAATCGTATTTTGATTCTTCTTAAATC
TCTCTCAATTTTCACTCTCTCTCTTTTCTTCTTATGTATTAGATTCTTTTACATAGCTA
ACACTTGTATAGAGAATTCAAAGTCTGGCTATTTTCGAAAGTTATCTTTCTCTTAAAA
AAAAAA

>G1811 Amino Acid Sequence (domain in AA coordinates: TBD)
MSINQYSSDPHYHSLMWQQQQQQHQNVDVVEEKEALFEKPLTPSDVGKLNRLVLPKQHA
ERYFPLAAAAADAVEKGLLLCFEDEEGKPWRFRYSYWNSSQSYVLTKGWSRYVKEKHLDA
GDVVLFHRHRSDDGRFFIGWRRRGDSSSSSDSYRHVQSNASLQYYPHAGAAVESQSGNS
KTLRLFGVNMECQLDSDWSEPTPDGSNTYTTNHDQFHYPQQQHYPPPYMDISFTGDM
NRTS*

>G182 (74..1366)

CGTCGACGATCAGATTCTTGGCTATAGCTGTATATATACACCAAGATACACTCATCATCG
TCATATATAGATTATGTCAGCGTCTCTGAGCTTCTTGACATGGAAAACCTCCAAGGAGA
CTTAACCGACGTCGTACGAGGAATCGGAGGCCACGTGTTATACCGGAGACTCCTCCCTC
GAACATCTGGCCTCTTCTCTGTGCATCCAAACACCATCACCGTCAGATCTTAACATAAA
CCCCTTCGGAGATCCCTTTGTGAGCATGGACGATCCACTCCTCCAAGAACTAAACTCCAT
CACAACTCCGGCTATTTCTCACCGTAGGAGATAACAACAACAACATTCAACAACA
TGGTTTCTTGGTTCCAAAGGTATTTGAGGAGGATCATATAAAGAGTCAATGTAGTATCTT
CCCAAGAAATCCGGATCTCGCATAGTAACATCATCCACGATTCTTCTCCGTGTAATTCTCC
GGCCATGTCCGGCTCACGTTGTGCGAGCCGACGAGCCGCTCGCCGAGAGGCATCATCAA
CGTAGACACAAACAGTCTTAGAACTGTCTATTGGTTGATGGTACCACGTTCTCCTCGCA
GATTGAGATATCTTCCCTCGGAATCTAGGCCTTAAAGAAGGAAGAGTCAGGCAAAGAA
GGTGGTGTGATTCCGGCCCCGGCTGCAATGAACAGCCGATCAAGCGGAGAAGTGGTTCC
ATCGGATCTATGGGCTTGGCGTAAATACGGTCAAAAACCTATCAAAGGCTCTCCTTTTCC
AAGGGGTATTATAGATGCAGCAGCTCAAAAGGTTGTTGAGCAAGAAAGCAAGTCGAAAG
AAGCCGAACCGATCCAAACATGTTGGTGATTACATATACCTCCGAACATAACCATCCTTG
GCCCCATCAACGCAACGCTCTCGCCGGCTCCACAGCTCCTCCACCTCCTCCTCATCTAA
CCCTAATCCTTCCAAACCTCAACCGCAACGTAACCTCCTCATCCATTGGCTCCCAAAA
CACCATCTACTTGCTTCTCTCCACACTCCTCCTCCTACCTCTCATCCTCCGCCATCAA
AGATGAACGAGGGGACGATGGAGTTGGAAAACGTAGATGATGATGATAACCAGAT
TGCTCCATACAGACCGGAGCTTCTATGATCATCAGCACCACCAACGATGATTTCTTTGCAGA
TCTTGAAGAGCTAGAAGGAGATTCTCTAAGCATGTTGCTTTCTCATGGCTGTGGCGGCGA
CGGAAGGATAAAACGACCGCTCGGATGGGATCAGCAATTTCTTCGGGTGGTTCGGGAGA
TAATAATTATAATAATTACGACGACCAAGACTCAAGGTCGTTATAGTATAGTGTAAATTA
CAGGTAAACAAATTATATTAAATTAAGTTGAGCTTGTGAAAATGAAGATCATATGGTCTG
GTCAGGTTGGGGG

>G182 Amino Acid Sequence (conserved domain in AA coordinates:217-276)
MCSVSELLDMENFQGDLDVVRGIGGHVLSPEPTPSNIWPLPLSHPTSPSDLNINPFGD
PFVSMDDPLLQELNSITNSGYFSTVGDNNNNIHNNGFLVPKVFEEDHIKSQCSIFPRIR
ISHSNIHDSPPCNSPAMSAHVVAASPRGIINVDINSPRNCLLDGTTFFSSQIQIS
SPRNLGLKRRKSQAKKVVCIPAPAAMNSRSSGEVVPDLWAWRKYGQKPIKGSPPFRGY
RCSSSKGCSARKQVERSRDTPNMLVITYTSEHNHPWPIQRNALAGSTRSSSSSNPNPS
KPSTANVNSSSIGSQNTIYLPSTTPPPTLSSSAIKDERGDDMELENVDDDDNQAIPYR
PELHDHQHPDDFFADLEELEGLSMLLSHGCGDGDKTTASDGISNFFGWSGDNNYN
NYDDQDSRSL*

>G1835 (1..969)

ATGATTGGAACAAGCTTCCCCGAGGATCTTGATTGTGGCAACTTCTTTGACAACATGGAT
GATCTCATGGACTTTCCCGGTGGAGATATCGATGTCGGTTTCGGCATAGGTGACTCCGAC

TCTTTCCCTACCATCTGGACCACTCATCACGACACGTGGCCTGCCGCTTCTGATCCTCTC
TTCTCTTCCAACCAACTCTGATTCATCACCTGAGCTCTATGTTCCGTTTGAGGACATT
GTTAAGGTGGAAAGACCTCCAAGCTTTGTAGAGGAAACATTGGTTGAGAAGAAGGAAGAT
TCGTTTTTCGACAAACACTGATTCATCATCTTCTCATAGCCAATTCAGGAGCTCAAGTCCA
GTGTCGGTTCTCGAGAGCAGCTCCTCCTCGTCTCAAACCACCAACACAACCTCCCTTGTT
CTCCCTGGAAAGCAGGTCGTCACGCACAAAACGCCCTCGTCCACCTGTCCAGGATAAA
GATAGAGTCAAAGACAATGTGTGCGGTGGTGACTCGCGCTCATCATTAGAATACCGAAA
CAGTTTCTCTCTGATCACAACAAGATGATCAACAAGAAGAAGAAGAAGGCCAAGATT
ACTTCTTCTCTCTTCTCGTCCGGGATTGATCTTGAAGTCAATGGAACAACGTCGATTCTG
TATTCTTCAGAGCAATATCCGCTTAGGAAATGTATGCACTGTGAGGTACCAAGACTCCA
CAGTGGAGGCTTGGTCCAATGGGTCCAAAGACACTTTGCAATGCGTGCAGGTGTACGTTAC
AAATCAGGGAGGCTTTTCCCGGAGTACCGTCCAGCTGCTAGTCCAACATTTACTCCAGCT
CTTCACTCAAACCTCACACAAGAAAGTGGCTGAAATGAGAAACAAGAGATGCAGTGATGGT
AGCTACATAACCGAAGAGAATGATCTGCAAGGGCTGATTCCGAACAATGCCTACATTGGC
GTAGACTAA

>G1835 Amino Acid Sequence (domain in AA coordinates: 224-296)
MIGTSFPEDLDCGNFFDNMDDLDFPGGDIDVGFIGDSDSFPTIWTTHDTPAASDPL
FSSNTNSDSSPELYVPFEDIVKVERPPSFVEETLVEKKEDSFSTNTDSSSSHSQFRSSSP
VSVLESSSSSSQTNTTSLVLPKGHRPRTKRPRPPVQDKDRVKNVCGGDSRLIIRIPK
QFLSDHNKMINKKKKKKAKITSSSSSSGIDLEVNGNNVDSYSSEQYPLRKMHCETVTKTP
QWRLGPMGPKTLCNACGVRYKSGRLFPEYRPAASPTFTPALHSNSHKKVAEMRNKRCSDG
SYITEENDLQGLIPNNAYIGVD*

>G1836 (47..610)
ATAACAAGCCTAGAACACTAGAAACTTCAAAAAAGAAAAAATCTTATGGAGAACAACAA
CGGCAACAACCACTGCCACCGAAAGGTAACGAGCAACTGAAGAGTTTCTGGTCAAAAGA
GATGGAAGGTAACTTAGATTTCAAAAATCACGACCTTCTATAACTCGTATCAAGAAGAT
TATGAAGTATGATCCGGATGTGACTATGATAGCTAGTGAGGCTCCAATCCTCCTCTCGAA
AGCATGTGAGATGTTTATCATGGATCTCACGATGCGTTCGTGGCTCCATGCTCAGGAAAG
CAAACGAGTCACGCTACAGAAATCTAATGTGATGCCCGAGTGGCTCAAACCTGTTATCTT
TGATTTCTTGCTTGATGATGACATTGAGGTAAAGAGAGAGTCTGTTGCGCCGCTGCTGA
TCCTGTGGCCATGCCACCTATTGACGATGGAGAGCTGCCTCCAGGAATGGTAATTGGAAC
TCCTGTTTGTAGTCTTGGAAATCCACCAACCACAACCAAAATGCAGGCATGGCCTGG
AGCTTGGACCTCGGTGTCTGGTGAGGAGGAAGAAGCGCTGGGAAAAAGGAGGTGACGA
CGGAAACTAATAAGTGAATACGTTTTAGGGTATTTTCAAGGGAATATGTAGTAAATAGT
CATGGATC

>G1836 Amino Acid Sequence (domain in AA coordinates: 30-164)
MENNNNGNNQLPPKQNEQLKSFWSKEMEGNLDKFNHDLPITRIKKIMKYDPDVTMIASEAP
ILLSKACEMFIMDLTMRSLWHAQESKRVTLOKSNVDAVAQTVIFDILLDDDI EVKRESV
AAAADPVAMPPIIDDGELPPGMVIGTFVCCSLGIHQPPQMQAWPGAWTSVSGEEEEARGK
KGGDDGN*

>G1838 (132..1628)
TTCCTTGGCATTCTCTTTAGAACTTTCGTACAAAATGCAAAACCTGAACCTCTAAAGCTA
AAAAAAAAGATTAGAGACTGTAAGTCTTTTATCAGATTTTCACTAGGAAAAAAGTTAC
AATCTTTTTTGATGGCTCCTCCAATGACGAATTGCTTAACGTTTTCTCTGTACCAATGG
AGATGTTGAAATCAACTGATCAGTCTCACTTCTCTTCTTACGACGATTCTTCTACTC
CTTATCTCATCGATAACTTCTATGCTTTCAAAGAAGAAGCTGAGATAGAAGCTGCTGCTG
CTTCAATGGCGGATTCAACAACCTTATCTACTTTTTTCGATCATTTCTCAGACTCAGATT
CAAAGCTGGAAGATTCCTCGGTGATTCTTTGTCCGTTACTCTGATAACCAACAGAGA
CCCAAGACTCTTCTCTCTCTCACTTCTACGATCCACGTCACCGCACCGTTGCCGAAG
GAGTTACAGGGTTCTTCTCTGATCATCATCAGCCAGATTTCAAGACGATAAACTCGGGAC
CAGAAATCTTCGATGACTCAACAACCTTCCAACATCGGTGGTACTCATCTCTCCAGTCACG
TGGTGGAGTCATCAACGACGCGGAAGTTAGGGTTTAACGGTGATTGCACCACCACCGGAG
GAGTTTTGTCTCTAGGGGTTAACAACACATCAGATCAACCTTTGAGCTGTAACAATGGCG
AGAGAGGTGGAAACAGTAACAAGAAGAAAACAGTTTCTAAGAAGGAAACATCAGATGATT
CAAAGAAGAAGATTGTGGAACATTGGGACAAAGAACTTCAATTTATCGTGGAGTCACCC
GACATAGATGGACTGGAAGATACGAAGCGCATCTATGGGATAACAGCTGTAGGAGGGAAG
GTCAAGCCAGAAAAGGACGTCAAGTGTAAGTTAGGTGGATATGACAAGGAAGATAGAGCAG

CTAGAGCCTATGACTTGGCAGCTTTAAATACTGGGGTTCTACTGCTACTACAAATTTTC
 CGGTCTCGAGTTATTCAAAGAAGTTGAGGAAATGAATCACATGACCAAGCAAGAGTTTA
 TTGCATCTCTTAGGAGGAAAAGTAGCGGTTTTTCGAGAGGAGCTTCAATATATAGAGGTG
 TCACAAGGCATCATCAACAAGGTCGCTGGCAAGCAAGAATCGGCCGTGTCGAGGAAACA
 AAGATCTTTACCTCGGAACCTTTGCAACCGAAGAGGAAGCAGCAGAGGCTTATGACATTG
 CAGCCATAAAGTTTCAAGGAATCAACGCAGTAACCTTTGAGATGAACAGGTATGACA
 TTGAAGCTGTCTGAATAGTTCTTTACCTGTAGGAGGAGCAGCTGCGAAACGCCACAAAC
 TCAAACTCGCTCTTGAATCTCCTTCTTCATCATCTCTGACCATAACCTCCAACAACAAC
 AGTTGCTTCCGTCCTCTTCTCCCTCGGATCAAAACCTAACTCAATCCCATGTGGCATTTC
 CATTGAGCCTTCAGTTCTCTATTACCACCAGAATCTTTTCAGCATTATCCTTTGGTCT
 CTGACTCTACAATTCAAGCTCCTATGAACCAAGCTGAGTTTTTCTTGTGGCCTAACCAAGT
 CTTACTAAATCATTGTTTGGTTCTTGTCTTAGACTTCTATTACCGCCTAACCGATGAC
 CCGAGGCTTATCTTCTTGAATCTTGCTATAAGGATGAATCTTTCAAGTTCCTTTTAAAC
 TGTAGGCTAAGACAGAAGTAGAGGGGAGAAAAGTTGAAGAATCTGAACTTTTGGGGTCA
 ATTTTGTATTAATGTTTTTCTTTTGTCAAGGGTGGATTATCGGTTTTTACTTATTTTT
 TGAATGTAATCGGCCTATAACGGTATAACTCTGTTTCCATTATGAATATTTTTCTCAA
 TTGAAAAA

>G1838 Amino Acid Sequence (conserved domain in AA coordinates: 229-305, 330-400)

MAPPMTNCLTFLSPMEMLKSTDQSHFSSSYDDSTPYLIDNFYAFKEEAEIEAAAASMA
 DSTTLSTFFDHSQTQIPKLEDFLGDSFVRYSDNQTTQDSSSLTPFYDPRHRTVAEGVTG
 FFSDDHQPDKFTINSQPEIFDDSTTSNIGGTHLSSHVESSTAKLGFNGDCTTTGGVLS
 LGVNNSTSDQPLSCNNGERGNSNKKKTVSKKETSDDSKKKIVETLGQRTSIYRGVTRHRW
 TGRYEHLWDNSCRREGQARKGRQVYLGGYDKEDRAARAYDLAALKYWGSTATTNFPVSS
 YSKELEEMNHMTKQEFIASLRRKSSGFSRGASIRGVTRHHQQGRWQARIGRVAGNKDLY
 LGTFATEEEAAEAYDIAAIKFRGINAVTNFEMNRYDIEAVMNSSLPVGGAAAKRHLKLA
 LESPSSSSDHNLQOQLPSSSPSDQNPNSIPCGIPFEPVLYYHQNFFQHYPLVSDST
 IQAPMNQAEFFLWPNQSY*

>G1843 (51..653)

CAGACATCACAATCAAATTAGGTCAGAAGATTAGTCGGAGAAAACAGCCATGGGAAGAA
 GAAAAGTAGAGATCAAACGAATTGAGAACAAAAGCTCTCGACAGTTACTTTCTGTAAAC
 GACGAAATGGTCTCATGGAGAAAGCTCGTCAACTCTCAATCTTTGTGAATCCTCCGTCG
 CTCTTATCATCATCTCTGCCACCGGAAGACTCTACAGCTTCTCCTCAGGTGATAGCATGG
 CCAAGATCCTCAGTCGTATGAATTAGAACAGGCTGATGATCTTAAACCTTGGATCTAG
 AAGAAAAAACTCTTAATTATCTTTTCGCACAAGGAGTTGCTAGAAACAATCCAATGCAAGA
 TTGAAGAAGCGAAAAGCGATAATGTAAGTATAGATTGTCTAAAGTCCCTGGAAGAGCAGC
 TCAAGACTGCTCTGTCTGTAAGTAGAGCTAGGAAGACAGAACTAATGATGGAGCTTGTGA
 AGACCCATCAAGAGAAGGAGAAGCTGCTGAGAGAGGAGAACCAGAGTTTGAATAACCAGC
 TTATAAGATGGGGAAGATGAAGAAGTCTGTGGAAGCAGAGGATGCAAGAGCAATGTCAC
 CGGAAAGTAGCTCTGACAACAAGCCACCGGAGACTCTCCTGCTTCTCAAGTAACCACCAT
 CACCAACGACTGATTCGAAAAATAAAATTGTAATAATTATGATTGTAGTTTCATAAGGA
 AAGCTACATACTGTATGTTAAAAATCCTCTTCTCCCCCTGCTACGAAAAGTCATCCAA
 GGAGATGCATCAAATAAAGTAATTGATTTTTATTGTTA

>G1843 Amino Acid Sequence (domain in AA coordinates: 2-57)

MGRRKVEIKRIENKSSRQVTFCKRRNGLMEKARQLSILCESSVALIIISATGRLYSFSSG
 DSMAKILSRYLEQADDLKTLDLEKTLNLYLSHKELLETIQCKIEAKSDNVSIDCLKSL
 EEQLKTALS VTRARKTELMELVKTHQEKEKLLREENQSLTNQLIKMGKMKKSVEAEDAR
 AMSPESSSDNKPPELTLK*

>G1853 (1..1860)

ATGAGAGGTTCTTGGTACAAGAGTGTTCCTCTGTTTTTGGTCTCAGACCACGGATCAGA
 GGGTTGTATTCTTCATTGTTGGTGTGTGGCTCTAGTTACTATTTTAGCACCATTGACA
 TCTAAATCGTATGATCTTCGTCAAGTTCGACACTTGTGCCGAACATTTATAGTAAGTAT
 AGGAGGATAAAGGAGCAAGCTGCTGTTGATTATCTTGATCTGAGGTCTCTTCTTTAGGG
 GCTAGTTTAAAGAGTTTCTTTTGTGGTAAAGAAAGAGAAAGTTATGTGCCTTGTAT
 AACATAACTGGGAATTTGCTTGTGGGCTTCAAGAGGGTGAGGAGTTAGATCGACATTGC
 GAGTTTGAAAGAGAGAAGGAAAGATGTGTAGTTCTCCTCCGAGAGATTATAAATAACCA
 CTTAGGTGGCCACTTGGTAGAGATATCATATGGAGTGGGAACGTGAAGATTACCAAAGAC
 CAGTTTCTTCTTCAGGAACGTGACAACGAGGTTAATGTTGCTTGAAGAGAATCAAATA

ACCTTTCACTCGGAGGACGGCCTGGTCTTTGATGGGGTCAAAGACTATGCTCGTCAAATT
GCTGAGATGATAGGTTTAGGAAGTGATACTGAATTTGCTCAAGCGGGTGACGGACTGTG
TTAGACATTGGTTGCGGATTTGGTAGCTTTGGTGCTCATTTAGTGTCTTTGAAGCTGATG
CCTATATGTATTGCTGAGTATGAGGCAACTGGGAGCCAAGTTCAGTTAGCTCTAGAGAGA
GGCCTTCTGCAATGATTGGCAATTTCTTTTCAAACAGCTTCCTTATCCAGCACTGTCT
TTTGATATGGTCCATTGTGCTCAATGTGGCACTACTTGGGATATCAAAGATGCAATGCTA
CTTTTGGAAAGTGGATCGTGTCTGAAACCCGGGGGATACTTTGTTTTAACTTCTCCACA
AACAAAGCACAGGGAAACTTACCAGATACCAAGAAAACGAGCATCTCAACACGGGTGAAT
GAGTTATCTAAGAAAATCTGTTGGAGTCTAACAGCTCAGCAGGATGAGACGTTTCTTTGG
CAGAAAACCTCTGATTCAAGTTGCTATTCTTCTCGTTCGCAAGCTTCTATACCTCTTTGC
AAAGATGGAGATAGCGTTCCGTATTACCACCCATTGGTTCCATGTATAAGCGGAACCACG
AGTAAACGCTGGATTTCTATACAGAACAGGTCTGCTGTTGCAGGAACAACCTCTGCCGGG
CTTGAAATTCATGGTTTAAACCCGGAAGAATTCTTCGAGGATACACAAATATGGAGATCA
GCTCTGAAAACTATTGGTCTTTGCTTACACCTCTAATTTTCTCTGACCATCCGAAGAGA
CCCGGTGATGAGGATCCTCTCCCGCTTTCAACATGATACGCAATGTGATGGACATGCAT
GCTCGTTTGGGAATTTAAATGCCGCTTTACTCGACGAAGGAAAATCTGCTTGGGTAATG
AACGTCGTCACGTCATGCACGTAATACTCTTCTATCATACTTGATCGTGGTTTCGCC
GGTGTCTACATGACTGGTGTGAACCATTCCCGACATATCCTCGAACATATGACATGCTT
CATGCCAATGAATTTCTCACACATCTTAGCTCAGAACGATGCAGCCTAATGGACTTGTTC
TTGGAGATGGACCGGATTTCTTCGCCCTGAGGGATGGGTTGTTCTAAGCGACAAAGTGGGA
GTAATCGAGATGGCTCGAGCACTTCAGCTCGAGTGCCTGGGAAGCAAGAGTCATTGAT
CTTCAAGATGGTAGTGACCAAAGACTTCTCGTCTGTCAAAAACCATTCATCAAAAAATAA
>G1853 Amino Acid Sequence (domain in AA coordinates: entire protein)
MRGSWYKSVSSVFLRPRIRGLLFFIVGVVALVTILAPLTSNSYDSSSSSTLVPNIYSNY
RRIKEQAAVDYLDLRLSLGASLKEFPFCGKERESYVPCYNITGNLLAGLQEGEELDRHC
EFEREKERCVRPRDYKIPLRWPLGRDIWISGNVKITKDQFLSSGTVTTRLMLLEENQI
TFHSEDGLVFDGVKDYARQIAEMIGLSDEFAQAGVRTVLDIGCGFGSFGAHLVSLKLM
PICIAEYEATGSQVQLALERGLPAMIGNFFSKQLPYPALSFDMVHCAQCGTTWDIKDAML
LLEVDRLKPGGYFVLTSPTNKAQGNLPDTKKTSTISTRVNELSKKICWSLTAQQDETFLW
QKTSDDSSCYSSRSQASIPCKDGDSPYHPLVPCISGTTSKRWISIQNRSAVAGTTSAG
LEIHGLKPEEFFEDTQIWRSAKKNYWSLLTPLIFSDHPKRPGEDEPLPPFNMRNVMDMH
ARFGNLSAALLDEGKSAWVMNVVPNARNTLPILDRGFAGVLHDWCEPPTYPRTYDML
HANELLTHLSSERCSLMDLFLEMDRILRPEGWVVLSDKVGVIEMARALAAVRWEARVID
LQDGSQRLLVLCQKPFICK*
>G1855 (1..1902)
ATGGCGAAAGAGAACAGTGGTCATCATCACCAAACAGAAGCAAGAAGAAAGAACTAACT
TTGATTCTTGGTGTAAGTGGACTCTGCATTTTGTCTATGTTTATAGGTGCATGGCAAGCC
AATACCGTCCCATCTTCTATCTCGAAGCTCGGATGCGAGACGCAATCAAACCTTCTTCG
TCCTCTTCTCTTCTCATCTTCAGAGTCAGCTGAAGTATTTCAAAGCCATAATCAG
ATTGAGTTAAAGGAAACAAACCAACCATTAAGTACTTTGAACCATGTGAATTATCTCTC
AGTGAGTACACTCCTTGTGAAGACCGACAAAGAGGAAGAAGATTTCGATAGGAACATGATG
AAATATAGAGAAAGACATTGTCTGTAAAAGATGAGCTTCTTTATTGTTTGATTCTCTCT
CCACCAAACACTACAAGATTCCATTTAAATGGCCACAAAGTAGAGACTATGCTTGGTATGAC
AATATCCCTCACAAGGAACCTTAGTGTGAGAAAGCAGTTCAAACTGGATTCAAGTTGAA
GGTGACCGCTTTAGATTCCCTGGTGGTACTATGTTTCTCGTGGAGCTGATGCTTAT
ATCGATGATATTGCTAGGCTTATTCTCTTACTGATGGTGAATCAGAACAGCTATTGAC
ACTGGATGTGGTGTGCAAGTTTTGGTGCTTACCTCTTGAAGAGAGACATTATGGCTGTG
TCTTTTGCTCCAAGAGACACTCATGAAGCTCAGGTACAGTTTGCTTTAGAACGCGGAGTT
CCTGCGATAATCGGATATGGGATCAAGAAGACTTCTTATCCAGCTAGAGCTTTTGAT
CTTGCTCATTTCTCGTTGTTTGATCCCTTGGTTTAAATATGATGGTTTGTACCTTATG
GAGGTCGACCGGTTTAAAGACCGGGCGGTTACTGGATCCTCTCGGGACCACCGATTAAC
TGGAAACAGTACTGGAGAGGTGGGAGAGAACAGAGGAGGATTTGAAGAAAGAGCAAGAT
TCAATAGAAGATGTAGCAAAGAGTCTTGTGGAAGAAAGTAAGTGAAGAAAGGTAAGT
TCAATTTGGCAAAGCCTCTCAATCACATTGAGTGTAAGAAAGCTCAAACAAAACAATAAG
TCACCTCCGATATGCAGCTCAGATAACCGGATTCGCTTGGTACAAAGACTTGGAAACT
TGTATAACACCATACCAGAAACAAACAATCCAGATGATTGAGCAGGCGGTGCACTCGAG
GATTGGCCAGACCGGATTCGCGGTACCTCCAAGAATCATCAGAGGAAGTATACCAGAA

GATCTTAACTCCAATCCTCAAAGAGTACACGAGCCTGGATCACATAACCAAGAGGAGGGT
ATTGATGTAAATAACGCAACAATCACGAGCATGAATAG
>G1881 Amino Acid Sequence (domain in AA coordinates:5-28, 56-79)
MRILCDACESAAAI VFCAADEAALCCSCDEKVHKCNKLASRHLRVGLADPSNAPSCDICE
NAPAFFYCEIDGSSLCLQCDMVVHVGGKRTTHRRFLLLRQRIEFPDGKPNHADQLGLRCQK
ASSGRGQESNGDHDHNMIDLNSNPQRVHEPGSHNQEEGIDVNNANNHEHE*
>G1882 (1..1200)
ATGGTTTTTCTTCATTTCTTACTTATCCTGATCATTTCATCAAACCTGGCAACAACAACAT
CAACCAATCACAACCACCGTTGGATTACGGGAAATAACATCAACCAACAGTTTCTTCCT
CACCATCCCCTCCCACCGCAACAGCAACAACGCCTCCGCAGCTTCACCACAACAACGGT
AACGGCGGAGTCGTGTTCCTCGGTGGACCTGGCGGGTTAATCCGACCAGGTTTCGATGGCG
GAAAGAGCAAGGCTAGCCAACATACCATTTACCTGAAACAGCCTTGAAGTGTCCAAGATGT
GACTCAACTAACACCAAAATCTGTCTTCAACAACACTACAGTCTCACTCAACCTCGCCAC
TTCTGCAAAGCATGCCGTGTTACTGGACACGTGGCGGTGCTCTAAGGAGCGTTCCCGTC
GGTGGCGGTTGCCGTAGAAACAAAGAACCACCAACAGCAGCGGTGGAGGTGGCGGTAGC
ACCAGTAGCGGTAACAGCAAGTCACAAGACAGCGCCACGAGCAACGACCAATACCACCAC
CGAGCCATGGCTAACCAATCAGATGGGACCACCTTCTTCGTCTATCGTCTTAAGCTCGTTG
CTGTCTTCTTACAACGCGAGGGTTAATCCCGGACATGATCATAACAGCAATAACAACAAC
ATACTTGGACTTGGATCATCTTTGCCTCCTCTTAAGCTTATGCCTCCTTAGACTTCACA
GACAACTTCACCTTACATACGGTGCCGTTTCAGCTCCTTCTTATCATATAGGCGGTGGA
AGCAGTGGAGGAGCGCGGCTCTTTTAAACGGTTTTGACCAGTGGAGATTCCCGGCAACA
AACCAACTTCTTTAGGCGGTTTAGACCCGTTTGATCAACAACATCAAATGGAGCAGCAG
AATCCAGTTTACGGATTGGTTACCGGGTCGGGTGAGTATCGACCTAAGAACATTTTCCAT
AACCTTATCTCTCTTCTTCTGCTGCTTCATCAGCTATGGTTACAGCCACCGCGTCGCAA
TTAGCTTCAGTGAAAATGGAAGATAGTAACAATCAGCTCAACTTGTCTAGACAACTTTTT
GGAGACGAACAACAGCTCTGGAATATTATGCGCTGCTGCAGCATCCACCGCAGCTGCA
ACAAGTTCGTGGAGTGAAGTCTCTAATAATTCAGTTCTTCTTCTACTAGCAATATATAA
>G1882 Amino Acid Sequence (domain in AA coordinates:97-125)
MVFSSFTYPDPHSSNWQQHQPIITTVGFTGNNINQQFLPHHPLPPQQQTPPQLHHNNG
NGGVAVPPGGPGLIRPGSMAERARLANIPLPETALKCPRCDSTNTKFCYFNYSLTQPRH
FCKACRRYWRGGALRSVPVGGGCRNRKRTKNSSGGGGSTSSGNSKSQDSATSNQYHH
RAMANNQMGPPSSSSSLSSLLSYNAGLIPGHDHNSNNNNILGLGSSLPPLKLMPLDFT
DNFTLQYGAVSAPSYHIGGGSSGGAALLNGFDQWRFPATNQLPLGLDLPFDQHQHMEQQ
NPGYGLVTGSGQYRPKNIFHNLISSSSSASSAMVTATASQLASVKMEDSNNQLNLSRQLF
GDEQLWNIHGAATAAATAATSSWSEVSNNFSSSSTSNI*
>G1883 (1..1110)
ATGGACGCTACGAAGTGGACACAGGGTTTTCAAGAAATGATGAACGTTAAACCAATGGAG
CAGATCATGATTCTTAATAACAACACACATCAACCAAAACACCACATCCAATGCAAGGCCA
AACACCATTTCTACATCTAACGGCGTCTCAACTGCTGGAGCAACCGTCTCCGGCGTAAGC
AACAAACAATAACAATACGCGCGTGTGTGGCGGAGAGGAAGCAAGACCACAAGAGAAACTA
AATTGTCCAAGATGCAACTCAACCAACACAAAGTTTTGTTACTACAACAATATAGTCTC
ACACAACCAAGATACTTCTGCAAGGTTGTGCAAGGTATTGGACCGAAGGTGGATCTCTT
AGGAATGTTCTGTGGGAGGAAGCTCAAGAAAGAACAGAGATCATCTTCATCTTCTTCA
TCAAACATCCTTCAGACAATACCATCTTCACCTCCAGATCTAAACCCGCCAATACTCTTC
TCAAACCAAAATCCATAATAAATCGAAAGGGTCATCACAAGATCTCAACTTGTGTCTTTC
CCAGTCATGCAAGATCAACATCATCATGTCATATGTCTCAGTTTCTTCAGATGCCT
AAGATGGAGGGAAAGTTGTAACATAACTCATCAGCAGCAGCCTTCATCATCTTCTTCTGTC
TATGTTCTCTCGTCTCTCTCTGTTTCAGCTCTTGAACCTTTAAGAACCGGAGTTAATGTT
TCTTCAAGATCAGGGATTAACTCATCGTTTCATGCTTCCGGTTCAATGATGGATTCAAAC
ACTGTGCTTTACACTTCTTCAGGGTTTCCAACAATGGTGGATTACAAGCCAAGTAATCTC
TCCTTCTCTACCGATCATCAAGGGCTTGGACACAATAGCAACAATAGGTCTGAAGCTCTT
CATAGTGATCATCACCACAAGGTAGAGTTTTGTTTCCATTTGGGGATCAAATGAAGGAG
CTTTTCATCAAGCATAACACAAGAGTTGATCATGATGATAATCAACAACAGAAGAGTCAT
GGAAATAATAATAATAATAAATCAAGCCCTAATAATGGATATTGGAGTGGGATGTTT
AGTACTACAGGAGGAGGATCTTCATGGTGA
>G1883 Amino Acid Sequence (domain in aa coordinates: 82-124)
MDATKWTQGFQEMMNVKPMQIMIPNNNTHQPNNTTSNARPNTILTSNGVSTAGATVSGVS

NNNNNTAVVAERKARPQEKLNCPRCNSTNTKFCYNNYSLTQPRYFCKGCRRYWTEGGSL
RNPVVGSSSRKNNRSSSSSSSSNIIQTIPSSLPDLNPPILFSNQIHNKSKGSSQDLNLLSF
PVMQDQH HHVHMSQFLQMPKMEGNGNITHQQQPSSSSSVYGSSSPVSALELLRTGVNV
SSRSGINSSFMPSGSMDSNTVLVTSSGFPTMVDYKPSNLSFSTDHQLGHNSNNRSEAL
HSDHHQQGRVLFPPFDQMKESSSITQEVDDHNNQQQKSHGNNNNNNSSPNNGYWSGMF
STTGGGSSW*

>G1884 (1..741)

ATGATGACGTCATCCCATCAGAGCAACACCACCGGCTTTAAACCGCGGCGGATCAAGACG
ACGGCGAAGCCACCACGTCAGATCAATAACAAAGAACCATCTCCGGCGACGACGCGGTG
CTCAAGTGTCCGAGATGTGATTCAAGTCAACACCAAATCTGCTACTACAACAACCTACAGC
TTGTCTCAGCCACGTCCTACTGCAAGAACTGTCGTCGTTACTGGACACGTGGCGGCGCC
CTCCGTAACGTTCCCATCGGTGGCTCCACTCGAAACAAGAACAAGCCTTGACAGCTCCAA
GTCATCTCTTCTCTCTCTTGTCTCGAACGGGACGTCATCGGCGTCTCGTGAGCTTGTA
AGAAACCATCCATCGACGGCAATGATGATGAGTCTCGGTGGATTCTCCGGCTATATG
TTCCGTTGGATCCTAACTTCAACCTTGCCTCGTCTTCTATCGAGTCTTTGAGTTCTTTT
AACCAAGATTTGCACCAAGCTTCAGCAACAAAGACTCGTCACTTCCATGTTTCTCCAA
GATTCTCTTCCGGTTAACGAGAAAACGGTTATGTTTCAGAACGTAGAGTTGATTCTCTCT
TCGACGGTGACGACGATTGGGTTTTCGATAGGTTGCCACTGGAGGAGGTGCAACAAGT
GGCAATCATGAAGATAATGATGATGGGGAGGGTAATTTGGGAAATTGGTTCCATAATGCT
ATAATAATGCTCTGCTCTAA

>G1884 Amino Acid Sequence (domain in AA coordinates: 43-71)

MMTSSHQSNNTTGFKPRRIKTTAKPPRQINNKEPSPATQPVLCPCRDVNTKFCYNNYS
LSQPRHYCKNCRRYWTRGGALRNVPIGGSTRNKNKPCSLQVISSPPLFSNGTSSASRELV
RNHPSTAMMMSSGGFGSYMFPLDPNFNLAASSIESLSFNQDLHQLQQORLVTSMFLQ
DSLVPNEKTVMFQNVLEIPPSTVTTDWFDRFATGGGATSGNHEDNDGEGNLGNWFHNA
NNNALL*

>G1891 (1..750)

ATGGATAACTTGAATGTTTTCGCAAATGAAGACAATCAAGTGAATGATGTGAAGCCCCCA
CCACCACCACCTCGAGTGTGTGCAAGGTGTGATTCTGATAATACTAAATTTTGTTATTAC
AACAACTACTGTGAGTTTCAGCCACGATACTTCTGCAAGAACTGTCGTAGATACTGGACT
CATGGTGGGGCTTTAAGAAACATACCAATTGGTGGAGTAGTCGTCCAAACGGGCAAGG
GTAAATCAACCTTCGGTTGCTCGGATGGTTTCTGTTGAGACCAACGAGGTAACAATCAA
CCTTCTCTAATGTTCAAGAAAACGTTTCATCTTGTGGATCTTTTGGTGCTTCATCTTCA
TCTTCTGTTGGTGCTGTTGGGAACCTTTTGGTTCTTGTATGATATTTCATGGTGGTATG
GTAAACAAATTTGCATCCAACCTCGAAGTTCGACCAAAATCATCGCTTAGCTTTCCATGAT
GGATCATTTGAGCAAGACTATTACGATGTTGGGTCCGATAATCTTTGGTCAACCAACAA
GTTGGTGGCTACGGTTATCACATGAATCCAGTGGATCAATTCAAGTGGAAACCAGAGCTTC
AACAACTATGAACATGAATTATAATAACGATAGCACTAGTGGAAGTAGCAGAGGATCT
GACATGAATGTGAACCATGATAACAAGAAGATCAGATACCGCAACTCTGTGATTATGCAT
CCTTGTCATCTGGAGAAGGATGGTCCTTGA

>G1891 Amino Acid Sequence (domain in aa coordinates: 27-69)

MDNLNVFANEDNQVNDVKPPPPPPRVCARCDSNTKFCYNNYCEFPQPRYFCKNCRRYWT
HGGALRNIPIGGSSRAKRARVNQPSVARMVSVETQRGNNQPFQSNQENVHLVGSFGASSS
SSVGAVGNLFGSLYDIHGMVITNLHPTRTVRPNHRLAFHDGSFEQDYDVGSDNLLVNQQ
VGGYGYHMPVDQFKWNQSFNNTMNMNYNDSTSGSSRGSMDMNVNHDNKKIRYRNSVIMH
PCHLEKDG*

>G1896 (1..951)

ATGTCCTCCCATAGCAATCTCCCCTCTCCCAAACAGTTTCTAAACCAGATCACCGTATC
TCCGGTACATCCCAAACCAAGAAACCACCGTCTTCTCCGTAGCTCAAGACCAACAAAC
CTAAATGCCCTCGTTGCAACTCTCCAAACACAAAGTTCTGTTACTACAACAACCTACAGT
CTCTCTCAACCTCGTCACTTCTGCAATCTTGTGCGCGTTACTGGACACGTGGCGGTGCT
CTAAGAAACGTCCCCTCGGTGGTGGTTGCCGGAAAACCAAAAATCTATCAAACCTAAT
TCCTCCATGAACACACTTCTTCTGCTCTTCTCTCTCAGAGGTTCTTCTCATCAATCATG
GAAGATTATCCAAATCTTCCCTCTCCGACAACAATGGATTTTCAGCTGGCCGGATTA
TCTCTCAACAAATGAACGATCTTCAACTTTTGAATAACCAAGAGTTCTTGATCTTAGG
CCCATGATGTCCTCGGGCGAGAAAACACACCCGTTGATGTCGGGTGGGTTTATCCCTA
ATGGGTTTTGGAGATTTCAACAACAACCATTCACCGACGGGTTCAACAACCGCCGGAGCA

AGCGACGGAACTTAGCTTCTTCTATAGAGACTTTGAGTTGTTTAAACCAAGATTACAC
TGGAGGCTTCAGCAACAGAGGATGGCGATGCTTTTTGGTAATTCTAAGGAAGAACTGTT
GTCGTCGAGAGGCCACAACCTATTCTTTATCGGAATCTTGAGATCGTAAACTCATCATCG
CCGTCGTCGCCGACGAAGAAAGGAGATAATCAGACAGAGTGGTATTTTGGTAATAACAGT
GATAATGAAGGAGTGATTAGTAATAATGCTAATACAGGAGGAGGAGGAAGTGAATGGAAC
AATGGAATTCAAGCTTGGACTGATCTTAATCATTATAATGCATTGCCTTGA

>G1896 Amino Acid Sequence (domain in aa coordinates: 43-85)
MSSHTNLPSPKPVKPDHRISGTSQTKPPSSSSVAQDQQLKPCRCNSPNTKFCYNNYS
LSQPRHFCKSCRRYWTRGGALRNVPVIGGCRKTKKSIKPNSSMNTLPSSSSSQRFFSSIM
EDSSKFFPPPTTMDFQLAGLSLNKMNDLQLLNNQEVLDLRPMSSGRENTFVDVGSGLSL
MGFGDFNNHNSPTGFTTAGASDGNLASSIETLSCLNQDLHWRLQQQRMAMLFGNSKEETV
VVERPQPILYRNLEIVNSSSPSPTKKGDNDQTEWYFGNNSDNEGVISNNANTGGGGSEWN
NGIQAWTDLNHYNALP*

>G1898 (1..630)

ATGCCGTCGGAACCAACCAACCCGACCCACCAGAGTTCAGCCCTCAACGCGCGCTTAC
CCACCGCCAAATCTGGCTGAGCCTCTTCCTTGTCTCGCTGCAACTCCACCACCACCAAG
TTCTGTTACTACAACAATAACCTCGCTCAGCCTCGCTACTACTGCAAATCTTGCCGC
CGTTACTGGACTCAAGGTGGTACACTCCGTCGCTCCCGTCGGTGGTGGAATCGTCGA
AGCTCCTCAAAACGTCACCGTTCTTTCTCCACCACCTGCCACCTCCTCTCTCTCTTCT
TCCGTATCACCACACGACACAAGAACCAGCCACGACTGAAGCGAGTCAAACCTAAGGTT
ACTAATTTAATTCAGGTCATGGAAGCTTTGCTTCTCTGTTAGGTTTAGGAAGTGGAAT
GGTGGGTTGGATTACGGGTTTGGGTACGGGTACGGGCTTGAGGAGATGAGTATTGGGTAT
CTTGAGATTCTTCCGTAGGAGAGATTCCGGTGGTTGATGGTTGTGGTGGTGACACGTGG
CAGATTGGGGAGATTGAAGGTAAAGTGGAGGAGACAGTTTGATATGGCCTGGTCTTGAG
ATCTCAATGCAAACCAACGATGTTAAGTGA

>G1898 Amino Acid Sequence (domain in AA coordinates: 31-59)
MPSEPNQTRPVRVQSTAAAYPPPNLAEPLPCPRCNSTTKFCYNNYNLAQPRYCKSCR
RYWTQGGTLRDVPVGGGTRRSSSKRHSFSTTATSSSSSSSVITTTTQEPATTEASQTKV
TNLISGHGSFASLLGLSGNGGLDYGFYGYGLEMSIGYLGDSVGEIPVVDGCGGDTW
QIGEIEGKSGGDSLIWPGLEISMQTNDVK*

>G1902 (1..615)

ATGCAGGATCCAGCAGCATATTACCAGACGATGATGGCGAAGCAACAACAACAACA
CCACAGTTTGCAGAGCAAGAACAGTTAAAGTGTCTCGTTGTGACTACCAAAACACTAAA
TTCTGTTACTACAACAATACTCTCACAGCCTCGTCACTTTTGCAAAAGCTGTCTGT
CGTTACTGGACTAAAGGCGGCGCTCTCCGTAACGTTCCGTCGGTGGTGGTTCTCGTAAG
AACGCAACCAACGATCCACTTCTTCTTCTTCTTCTGCTTCTCTCTTCCAACAGTAGC
CAAAACAAGAAGACGAAAAACCCGGATCCGGATCCTGATCCACGTAATTTCAAAAACCG
GATTGGATCCGACCCGGATGCTTTACGGGTTTCCGATCGGTGACCAAGCGTGAAGGGT
ATGGAGATTGGTGAAGCTTTAGCTCGTTGTTGGCGAATAATATGCAGCTTGGTCTTGA
GGAGGAGGGATCATGCTTGACGGGTGGGTTGGGATCATCCGGGTATGGGTTTGGGTTT
AGGAGAACCGAACCGGTAATAATAATAATAACCCATGGACCGATCTGGCTATGAACAGA
GCGGAGAAAACTGA

>G1902 Amino Acid Sequence (domain in AA coordinates:31-59)
MQDPAAYYQTMMAKQQQQQQPQFAEQEQLKCPRCDSPTNTKFCYNNYNLSQPRHFCKSCR
RYWTKGGALRNVPVGGGSRKNATKRSTSSSSSSASSPSNSSQNKKTKNPDPPDPRNSQKP
DLDPTRMLYGFPIGDQDVKGMEIGGSFSSLLANNMQLGLGGGGIMLDGSGWDHPPMGLGL
RRTEPGNNNNPNWTDLAMNRAEKN*

>G1904 (1..924)

ATGCAAGATATTCATGATTTCTCCATGAACGGAGTTGGTGGTGGGGGAGGAGGAGGG
AGGTTTTTTCGGTGGAGGAATCGGCGGCGGAGGAGGTGGTGATCGAAGGATGAGAGCTCAT
CAGAACATATACTTAACCATCATCAATCTCTCAAGTGTCTCGTTGTAATTCTCTTAAC
ACAAAGTTCTGTACTACAACAATTACAATCTTCTCAGCCTCGTCACTTTTGCAAGAAC
TGTCGTCGTTACTGGACTAAAGGTGGTGTCTCCGTAACGTTCCCGTCGGAGGTGGTTGC
CGGAAAGCTAAACGTTTCAAAAACAAACAGGTTCCGTCGTCGTCATCAGCCGACAAACCA
ACGACGACGCAAGATGATCATCAGTGAGGAGAAATCGAGTACAGGATCTCACTCTAGC
AGCGAGAGCTCTTCTCTCACCGCTTCTAACTCTACCACCGTCGCCCGCTCTCCGTCACC
GCGGCGGCGGAAGTTGCTTCGTCGGTTATTCAGGTTTTGATATGCCTAATATGAAAATT

TACGGTAACGGGATCGAGTGGTCGACGTTACTTGGACAAGGCTCATCGCCGGTGGTGTT
 TTCTCGGAGATCGGTGGTTTTCCGGCGGTTTCAGCTATTGAACTACACCGTTTGGATT
 GGGGGTAAATTCGTAAATCAAGATGATCATCTGAAGTTAGAAGGTGAACTGTACAGCAG
 CAACAGTTTGGAGATCGAACGGCTCAGGTTGAGTTTCAAGGAAGATCTTCGGATCCGAAT
 ATGGGATTGAAACCGTTGGATTGGGGAAGTGGCGGTGGAGATCAAACACTGTTTGATT
 ACCAGTACCGTTGATCATGCATACGGAGTCAAAGTCAATGGACGTCGTCTGACCAAGAT
 CAGAGTGGTCTCTACCTTCCTTGA

>G1904 Amino Acid Sequence (domain in aa coordinates: 53-95)
 MQDIHDFSMNGVGGGGGGGRFFGGGIGGGGGGDRMRHQNILNHHQSLKCPRCNSLN
 TKFCYNNYNNLSQPRHFCKNCRRYWTKGGVLRNVPVGGGCRKAKRSKTKQVPSSSSADKP
 TTTQDDHHVEEKSSSTGSHSSSESSSLTASNSTTVAASVTAAAEVASSVIPGFDMPNMKI
 YGNGIEWSTLLGQSSAGGVFSEIGGFPAVSAIETTFPGFVGKFPVNQDDHLKLEGETVQQ
 QQFGDRTAQVEFQGRSSDPNMGFEPLDWGSGGGDQTLFDLTSTVDHAYWSQSQWTSDDQD
 QSGLYLP*

>G1906 (1..795)
 ATGGTGAACGTGCTCGGATCGCAAAAGTCCCATTGCCTGAAGCAGCTCTAAATTGCCCT
 AGATGTGACTCAACCAATACTAAGTTCTGTTACTTCAATACTATAGCCTTACTCAACCT
 CGCCATTTCTGCAAAACATGTCGTCGCTATTGGACACGTGGCGGTTCTTGAGGAATGTT
 CCTGTTGGAGGAGGCTTTAGGAGGAACAAGAGAAGCAAATCCAGATCGAAATCTACGGTC
 GTGGTCTCGACTGATAATACTACTAGTACTTCATCACTTACTTCTCGCCCAAGTTACTCA
 AACCCTAGCAAGTTTCATAGCTACGGTCAAATCCCGGAGTTTAAATCCAACCTTGCCCATC
 TTGCCTCTCTCCAAAGCCTTGGAGATTACAATTCAAGCAACACTGGATTAGATTTTGGT
 GGAACCTCAAATAAGCAACATGATAAGTGGTATGAGTTCTAGTGGTGGGATCTTGGATGCA
 TGGAGAATACCTCCATCACAACAAGCTCAGCAATTCCTTTCTTGATCAACACTACCGGA
 TTGGTGCAATCTTCAAACGCGTTATATCCATTACTAGAAAGGCGGGTTAGCGCCACGCAA
 ACAAGAAATGTGAAGCGGAAGAGAATGATCAGGATCGGGGTAGGGATGGGGATGGAGTG
 AATAACTTATCAAGAACTTTTTGGGTAATATCAACATAAACTCAGGCAGGAACGAGGAA
 TACACATCATGGGGAGGTAACAGTTCTTGGACCGGTTTACCTCCAACAACCTCAACAGGC
 CATCTCTCATTCTAA

>G1906 Amino Acid Sequence (domain in AA coordinates: 19-47)
 MVERARIAKVLPEAALNCPKRDSTNTKFCYFNYSLTQPRHFCKTCRRYWTRGGSRLNV
 PVGGGFRNRKRSKRSKSTVVVSTDNTTSTSSLSRPSYSNPSKFHSYQIPEFNSNLPI
 LPPLQSLGDTNNSNTGLDFGGTQISNMISGMSSSGILDARIIPPSQQAQQFPFLINTTG
 LVQSSNALYPLLEGGVSATQTRNVKAEENDQDRGRDGDGVNNLSRNLGNININSGRNEE
 YTSWGGNSSWTGFTSNNSTGHLSF*

>G1913 (1..744)
 ATGGAGAGAGCAGAGGCCTTGACATCATCGTTTATATGGCGGCCAAACGCAAACGCAAAC
 GCGGAGATCAGCCGAGTTGTCCAAGATGTGGATCCTCTAACACAAAGTTCTGTTACTAC
 AACAACTATAGCCTCACTCAGCCTCGCTACTTCTGCAAAGGCTGCCGAGATATTGGACC
 AAAGTGGTTCCCTCCGCAATGTTCTGTAGGCGGTGGCTGTGCGAAATCCCGCCGCCCC
 AAATCATCTTCTGGTAACAATACTAAAACTAGCCTAACCGCTAATCTTGGCAACCCCGGT
 GGTGGTTCACCAAGCATCGATCTTGCTCTTGTTTACGCCAATTTCTTGAATCCAAAGCCT
 GACGAATCTATACTACAAGAAAATTGCGACTTAGCCACTACGGATTTTTTGGTAGATAAT
 CCTACCGGCACTTCCATGGACCCTTCATGGAGTATGGACATCAATGATGGTCATCATGAT
 CATTATATTAATCCGGTGGAAACACATTGTGGAGGAATGTGGTTATAATGGCTTGCCCTCA
 TTTCTGTGTAAGAGCTTCTCTTTAGACACTAATGGTGTGTTGGTCTGATGCTTTGTTG
 ATTGGTCATAACCATGTAGACGTTGGCGTGAATCCCGTTTCAAGCTGTACACGAACCGGTG
 GTTCATTTCTGCTGAAGAATCCAATGATTCCACCAATCTCTTGTGTTGGAAGTTGGAGCCCT
 TTTGATTTCACTGCCGATGGATGA

>G1913 Amino Acid Sequence (domain in AA coordinates: 27-55)
 MERAEALTSSFIWRPNANANAEITPSCPRCGSSNTKFCYNNYSLTQPRYFCKGCRRYWT
 KGGSLRNVVGGGCRKSRPKSSSGNNTKTSLTANSNPGGGSPSIDLALVYANFLNPKP
 DESILQENCDLATDFLVDNPTGTSMDPSWSMDINDGHHHDHYINPVEHIVEECGYNGLPP
 FPGEELSLDINGVWSDALLIGHNVHDVGVTVPVQAVHEPVVHFADESNDSTNLLFGSWSP
 FDFADG*

>G1914 (1..945)
 ATGGAGAGATACAAGTGTAGATTTTGCTTCAAGAGCTTCATCAATGGAAGAGCTTTAGGT

BNSDOCID: <WO_ 03013227A2 1A>

L*

>G1930 (76..1077)

ATTCACATTACTAATCTCTCAAGATTTTCAACAATTTCTTGTGATTTTCTCTCAGTTTCTT
ATTTTCGTTTCATAACATGGATGCCATGAGTAGCGTAGACGAGAGCTCTACAACCTACAGAT
TCCATTCCGGCGAGAAAGTCATCGTCTCCGGCGAGTTTACTATATAGAATGGGAAGCGGA
ACAAGCGTGGTACTTGATTGAGAGAACGGTGTCTGAAGTCTGAAGTCTGAAGCCGAATCAAGA
AAGCTTCTCTTCTCAAGATTTCAAGGTGTTGTTCTTCAACCAATGGAAGATGGGGAGCT
CAGATTTACGAGAAACATCAACGCGTGTGGCTTGGTACTTTCAACGAGGAAGACGAAGCA
GCTCGTCTTACGACGTCGCGGCTCACCGTTTCCGTGGCCGCGATGCCGTTACTAATTTT
AAAGACACGACGTTCTGAAGAAGAGGTTGAGTTCTTAAACGCGCATTCGAAATCAGAGATC
GTAGATATGTTGAGAAAACACACTTACAAAGAAGAGTTAGACCAAGGAAACGTAACCGT
GACGGTAACGGAAAAGAGACGACGGCGTTTGCTTTGGCTTCGATGGTGGTTATGACGGGG
TTTAAACGCGCGGAGTTACTGTTTGAAGAAAACGGTAACGCCAAGTGACGTCGGGAAACTA
AACCCTTTAGTTATACCAAAACACCAAGCGGAGAAACATTTTCCGTTACCGTTAGGTAAT
AATAACGTTCTCCGTTAAAGGTATGCTGTTGAATTTTCAAGACGTTAACGGGAAAGTGTGG
AGGTTCCGTTACTCTTATTGGAATAGTAGTCAAAGTTATGTGTTGACCAAGGTTGGAGT
AGATTCGTTAAAGAGAAGAGACTTTTGTGCTGGTGAATTTGATCAGTTTTTAAAGATCCAAC
GATCAAGATCAAAAATTCTTTATCGGGTGGAAATCGAAATCCGGGTTGGATCTAGAGACG
GGTCGGGTTATGAGATTGTTTGGGGTTGATATTTCTTTAAACGCCGTCGTTGTAGTGAAG
GAAACAACGGAGGTGTTAATGTCGTCGTTAAGGTGTAAGAAGCAACGAGTTTTGTAATAA
CAATTTAACAACTTGGGAAAGAAAAAAGCTTTTTGATTTTAATTTCTCTTCAACGTTA
ATCTTGCTGAGATTA

>G1930 Amino Acid Sequence (domain in AA coordinates: 59-124)

MDAMSSVDESSTTTDSIPARKSSSPASLLYRMGSGTSVVLDSENGVEVEVEAESRKLPS
RFKGVVPQPNRWWGQAIYEKHQRVWLGTFFNEEDEAARAYDVAAHRFRGRDAVTNFKDTTF
EEVEEFLNAHSKSEIVDMLRKHTYKEELDQRKRNRDNGKETTAALASMVMTGFKTAE
LLFEKTVTPSDVGKLNRLVIPKHQAQKHFLPLGNNNVSVKGMLLNFEDVNGKVWRFRYS
YWNSSQSYVLTKGWSRFVKEKRLCAGDLISFKRSNDQDQKFFIGWKSksGLDLETGRVMR
LFGVDISLNAVVVVKETTEVLMSSLRCKKQRL*

>G195 (51..1031)

TTTTCTTTTCTTTCTTTTGGTTTAAAGTTTTTCTCTTTGTTCTTCGTCATGTCTCATG
AAATCAAAGATCTTAACAACTATCACTACACTTCATCGTATAATCATTACAATATCAACA
ACCAAAATATGATTAATCTCCCTTACGTTTCTGGTCCATCTGCTTATAATGCAAACATGA
TCTCATCATCACAAAGTAGGTTTTGATCTACCCCTCGAAGAACTTGAGTCCTCAAGGAGCCT
TCGAGTTGGGTTTTCGAGCTTTCTCCATCTCTTCTGACTTTTTTAATCCTTCCCTCGATC
AAGAGAACGGTTTGTATAATGCTTATAATTATAATAGTAGTCAAAAGAGTCATGAAGTTG
TCGGTGATGGTTGTGCAACCATTAAAGAGTGAAGTTAGGGTTTCAGCATCTCCTTCTTCAA
GTGAGGCCGATCATCATCCAGGAGAAGATTCCGGCAAGATCCGGAAGAAAAGAGAAGTTC
CGGATGGAGGAGAAGATGATCAACGCTCTCAGAAAGTAGTTAAAACAAAGAAGAAAGAGG
AGAAGAAAAAAGAGCCACGAGTCTCGTTTCATGACTAAGACCGAAGTTGATCATCTCGAAG
ACGGCTATCGTTGGAGAAAGTATGGCCAAAAAGCAGTCAAAAACAGTCCTTATCCGAGGA
GTTACTATAGATGCACGACTCAGAAGTGCAACGTGAAGAAGAGAGTGGAGAGATCTTACC
AAGACCCAACGGTCGTCATCACAACTACGAGAGTCAACACAACCATCCGATCCCGACCA
ATCGTCGGACAGCAATGTTCTCTGGAACACCGCATCTGATTATAACCCATCATCGTCTC
CAATATTCTCCGATCTCATCATCAATACTCCAAGAAGCTTCTCAAATGATGATCTCTTCC
GTGTGCCATACGCTAGTGTGAACGTGAACCTAGTTATCATCAACAGCAACATGGATTTC
ATCAACAGGAGAGTGAGTTCGAGCTCTTGAAGGAGATGTTTCTTCGGTTTTCTTCAAAC
AAGAGCCTTGATGATATAATATAATATAGAAACAATTTTTTTCTGCTAAGAAATATAGA
ACAAAACCTTGGATGCATAATAAGTGATGATAGTGTTATTTATTTTTCATGTATATATT
ATACATGTTTTTTGTTAACTAGCTATAGGATATACTGGTAGTAATTAAGCATAAATATGGAG
CCCTTCGACTTATTACAATAATTTTTGGTATGGAAAAANTTNGNTACATGCCTGCCTTTT
NNNTTNNNGG

>G195 Amino Acid Sequence (domain in AA coordinates: 183-239)

MSHEIKDLNNYHYTSSYNHYNINNNQNMNINLPVVSGPSAYNANMISSQVGFDLPSKNLSP
QGAFELGFELESPSSSDFFNPNSLDQENGLYNAYNINSSQKSHEVVGDGCATIKSEVRVSAS
PSSSEADHHPGEDSGKIRKKREVRDGGEDDQRSQKVVKTKKKEKKKEPRVSFMTKTEVD
HLEDGYRWRKYQKAVKNSPYPRSYRRTTQKCNVKKRVERSYQDPTVVITTYESQHNHP

IPTNRRRTAMFSGTTASDYNPSSSPIFSDLIINTPRSFSNDDLFRVPYASVNVNPSYHQQQ
HGFHQESEFELLKEMFPSVFFKQEP*

>G1954 (196..1440)

ATTTATGACTTCTCAATACAAAAAGCTCCCTCACTTTTTTAAGTTTTGTCTTCTCTAAT
CCGTCTTCTTCTACTATCTTGATGTCTTGCGTCTTTTATATACATCTCTCGTAAACCCCT
AGCAAATCATACAAGGTCAAGAAGCTTGACCTTCATTAGACTTAAGCAGTTTATAATCAA
CTACCACGAATAGCAATGGATAAAGATTACTCGGCACCAAACCTTCTTAGGTGAATCCTCA
GGCGGTAACGATGATAACAGCTCTGGTATGATAGACTATATGTTCAATAGAAACCTTCAA
CAACAACAAAAGCAATCGATGCCACAACAGCAGCAACATCAACTCTCTCCTTCCGGATTT
GGAGCAACACCCCTTGATAAAATGAACCTTCTCTGATGTGATGCAGTTTGGCGACTTCGGT
TCGAAACTTGCGTTGAACCAGACCAGAAACCAAGACGATCAAGAAACCGGGATTGACCCC
GTTTATTTCTTGAAGTTCCTGTCTTGAACGACAAAATAGAGGACCATAACCAAACCCAA
CATCTCATGCCCTTCTCATCAGACGTCTCAAGAAGGAGGTGAGTGTGGAGGAAACATAGGC
AATGTGTTTTCTTGAAGAAAAAGAAGATCAAGACGATGACAACGACAACAACCTCCGTGCAA
CTACGTTTTTATTGGAGGAGAAGAAGATAGGGAGAACAGAATGTTACGAAAAAGGAG
GTGAAGAGCAAGAGGAAGAGAGCTAGAACGAGCAAGACCAGCGAAGAAGTGGAAAGCCAA
CGGATGACTCATATCGCGGTGCAAGAAACCGTAGGAAGCAATGAATGAGCATCTTCGT
GTCCTTAGATCTCTCATGCCCTGGCTCCTACGTTCAAAGGGGAGACCAAGCGTCAATCATA
GGAGGAGCAATAGAGTTTGTGAGAGAGCTCGAGCAACTCCTACAATGTCTTGAATCACAG
AAGCGTCGAAGAATCTTAGGAGAAACCGGTAGGGACATGACAACGACAACGACTTCTTCT
TCTTCTCCCATAACTACGGTAGCGAACCAAGCACAACCGCTCATTATTACGGGAAATGTA
ACCGAGCTAGAGGGCGGAGGAGGGCTTCGGGAGGAGACTGCGGAGAACAAAGTCGTGCTTG
GCTGACGTGGAGGTGAAGCTGCTAGGGTTTGACGCCATGATCAAGATACTTTCAAGAAGA
AGGCCGGGACAGCTGATTAAGACTATAGCTGCTTTGGAGGATCTTCATCTCTCTATTCTT
CACACTAACATCACTACCATGGAACAAACCGTCTCTACTCCTTTAATGTCAAGATAACA
AGTGAACGAGGTTTACGGCAGAAGACATAGCAAGTTCATCCAACAGATATTTAGTTTC
ATTTCATGCAAAATACCAACATATCTGGAAGCTCTAACCTGGGAAATATTGTGTTTACTTGA
AAATCATCACACGGCGACAACCTTTGTACTCTGGTGAAGATTACAGTACGTAATAATCTCT
ACATATTGGGTTTTATTCTCCAAGCATTTGGAAGAGTGTTTAAAGTTAAAGGGAGTGCTTA
CTTTATTTTTTTGGGGCTTTTTTTCATGCAATTTAAATTTTAGTGATGATTGTGTGCTTG
TAATGTTAGAACTCGTTGTTGTGATTCTGCTGCTTTGATTGTAGGTTTTGAACAAGCG
GTTTAGAATGCTAAACCACTTATTACTTGAAATAACTTTTTTCACAAAAA

AAGAAAAAA
>G1954 Amino Acid Sequence (domain in AA coordinates:187-259)
MDKDYSAPNFLGESSGNDNSSGMIDYMFNRNLQQQQKQSMPPQQQHQQLSPSGFGATPF
DKMNFSDVMQFADFGSKLALNQTRNQDDQETGIDPVYFLKFPVLNDKIEDHNQTOHLMPS
HQTSQEGGECGGNIGNVFLEEKEDQDDNDNNSVQLRFIGGEEEDRENKNVTKKEVKS
KRARTSKTSEEVESQRMTHIAVERNRRKQMNHLRLVRLSLMPGSYVQRGDQASIIIGGAIE
FVRELEQLQLCLESQKRRRILGETGRDMTTTTSSSSPITTVANQAQPLIITGNVTELEG
GGGLREETAENKSLADVEVKLLGFDAMIKILSRRRPGQLIKTIAALEDLHLSILHNTNIT
TMEQTVLYSFNVKITSETRTAEDIASSIQIFSIHANTNISGSSNLGNIVFT*

>G1958 (107..1336)

GTACCGTCGACCGATTATCCCCAAGAGGAGAATCCTCATAATCATTTTTCTCCGATTGAT
TCGTCTTCTTGGTCTTGATTGCTTCATGAATTTCTAGGACAACAATGGAGGCTCGTCC
AGTTCATAGATCAGGTTGAGAGACCTCACACGCACCTTCTCAATCCCATCTACAAAAA
ACCTTCACCAAGTAGAAGATAGTTTCATGAGATCAGATAACAACAGTCAGTTAATGTCTAG
ACCATTAGGACAAACCTACCATTTACTTTTCATCTAGTAACGGTGGAGCTGTTGGACATAT
ATGTTCTTCTTCATCATCTGTTTGTGCAACCAATCTCCATTACTCAACTATGGTATCTCA
TGAGAAAACAACAACACTACACAGGAAGCAGCAGTAATAATGCTGTGCAGACACCAAGCAA
CAACGATAGTGCTTGGTGTCTGATTTCATTGCCAGGAGGGTTTCTTGACTTCCATGAAC
CAACCCGGCGATTCAAAACAACCTGTCAGATTGAGGATGGTGGCATTGCGGCTGCTTTTGA
TGACATTCAAAACGAAGTGATTGGCATGAATGGGCTGACCATTGATCACTGATGATGA
TCCTTTGATGTCTACTAACTGGAATGATCTCTTGCTTGAAACAAATCCAATTCAGATTC
AAAGGACCAGAAGACACTGCAATTCGCAACCTCAGATTGTTTCAGCAGCAACCTTCTCC
GTCTGTGGAATTGCGACCTGTTAGCACAACATCTTCAACAGCAATAACGGAACGGGCAA
GGCAGCAATGCGTTGGACGCCAGAGCTTCACGAGGCTTTGTTGAGGCTGTCAACAGTCT
TGGCGGTAGTGAAAGAGCTACTCCTAAAGGGGTACTGAAGATTATGAAGTTGAAGGCTT

GACTATATATCATGTTAAAAGCCATTTACAGAAATATAGGACAGCTAGATATCGGCCAGA
ACCATCAGAAACTGGTTCGCCAGAAAGGAAGTTGACACCGCTTGAACATATAACATCTCT
TGATTTGAAAGGTGGGATAGGTATTACAGAGGCTCTACGACTTCAGATGGAAGTACAGAA
GCAACTCCATGAGCAGCTCGAGATTCAAAGAAACCTGCAACTCCGAATAGAAGAACAAGG
CAAGTACCTGCAAATGATGTTTCGAGAAGCAAACTCTGGTCTTACCAAAGGGACAGCCTC
AACATCAGATTCCGCAGCCAAATCTGAACAAGAAGACAAGAAGACTGCTGATTCGAAGGA
GGTTCAGAGAAGAAACCAGGAAATGTGAGGAACCTAGAAATCTCCACAGCCAAAGCGTCC
CAAATCGATAATTGAAAGTATTGGTCTTTTGTGCTGGATAATCTCGGAGTTTCAGAGTTAA
CAGTGATAGAGAGAACGAGCTCTTATCTTGAGGTTCTTCAGGACTTCTCTCGCGGCCGCT
CTAG

>G1958 Amino Acid Sequence (domain in AA coordinates: 230-278)

MEARPVHRSGSRDLTRTSSIPSTQKPSPVEDSFMRSDNNSQLMSRPLGQTYHLLSSSNGG
AVGHICSSSSSGFATNLHYSTMVSHEKQQHYTGSSSSNNAVQTPSNNSAWCHDSLPGGF
DFHETNPAIQNNCQIEDGGIAAFDDIQKRSDWHEWADHLITDDPLMSTNWNLDLLETN
SNSDSKDQKTLQIPQPIVQQQSPSPVELRPVSTTSSNSNNGTGKARMRWTPELHEAFVE
AVNSLGGSERATPKGVLLKIMKVEGLTIYHVKSHLQKYRTARYRPEPSETGSPERKLTPL
HITSLDLKGGIGITEALRLQMEVQKQLHEQLEIQRLQLRIEEQKYLQMMFEKQNSGLT
KGTASTSDSAKSEQEDKKTADSKEVPEEETRKCEELESPQPKRPKIDN*

>G196 (111..1421)

TCGACATCAGATTCTCTCACGGATTCTTAATCATTTTTATTATATTTGGATATTTGCTA
ATTCTTCCCGTGTATAAATCTCATATAAACACGCATCATACATATATATTATGTGCAGCG
TCTTTGAGTTTCAAGACATGGACAACCTCCAAGGAGATCTAACAGACGTCGTACGAGGAA
TAGGATCAGGCCACGTGTCAACATCTCCTGGACCACCGGAAGGTCCATCTCCGAGCAGCA
TGTCTCCGCCGCCAACATCAGATCTCCACGTGGAATTCCTCCGCCGCTACTTCTGCCA
GCTGTCTCGCAAATCCCTTCGGAGACCCGTTCTGTAAGCATGAAGGATCCTCTCATCCACC
TCCCGGCCAGCTACATCTCCGGCGCCGGTGATAATAAAGCAACAAAAGTTTTGCAATCT
TTCCAAAGATTTTTGAGGATGATCATATTAAGAGTCAATGCAGTGTCTTCCCAAGAAATA
AGATCTCGCAAAGTAACAATATCCACGATGCCTCCACGTGTAATCTCCGGCCATAACCG
TCTCCTCTGCCGCCGTAGCAGCTTCGCCGTGGGGCATGATCAACGTTAATACCACTAACA
GTCCAAGAAACTGTTTACTTGTGCGATAATAATAACAACACGTCATCATGCTCACAGGTT
AGATCTCTTCTTCCCTCGGAATCTCGGAATTAAGAGAAGGAAGAGCCAGGCAAAGAAAG
TGGTGTGCATACCGGCTCCAGCCGCTATGAACAGCCGGTCCAGTGAGAGAAGTTGTTCCGT
CTGATCTATGGCTTGGCGAAGTAGCAGCTCAAAAGGTTGTTTCTCCTTATCCAA
GGGTTTACTACAGATGTAGCAGCTCAAAAGGTTGTTTCTCAGCTAGGAAACAAGTGAACGTA
GCCGCACTGATCCAAACATGTTAGTCATTACTTACACCTCTGAGCATAACCACCCATGGC
CTACTCAACGCAACGCTCTCGCAGGTTCCACTCGTTCCTCTTCTCCTCCTCTTTAAACC
CTTCTTCCAAATCCTCAACCGCAGCCGCCACTACTTCTCCCTCATCCAGAGTTTTCCAAA
ACAACAGCAGCAAAGACGAACCCAATAACTCCAACCTTGCTTCTCTTCCACTCATCCTC
CTTTTGACGCCGCCGCAATTAAGGAGGAGAACGTGGAAGAGCGTCAGGAAAAGATGGAGT
TCGATTATAATGACGTTGAAAATACCTATAGACCGAGTTGTTGCAAGAGTTTCAACATC
AGCCGGAGGATTTCTTTGCCGATCTCGACGAGCTTGAGGGAGATTCTTTGACTATGTTGC
TCTCTCAGTAGCGCGGAGGCAACATGGAAAACAAAACGACGATTCCAGACGTTTTTA
GTGATTTCTTTGACGACGACGAGTCCTCAAGGTCGTTATAAATATTGTTGTTAATGTATA
CATAGAAATGAAATTATTCATGTAATTCGTTTTGTGTTAAATGACGGTATTTGCCTTTGC
A

>G196 Amino Acid Sequence (conserved domain in AA coordinates: 223-283)

MCSVFQMDNFQGDLDVVRGIGSGHVSPSPGPPEGSPSSMSPPPTSDLVHVEFPSAA
TSASCLANPFGDPFVSMKDP LIHL PASYISGAGDNKSNKSF AIFPKIFEDDH IKSQCSVF
PRIKISQSNNIHDASTCNSPAITVSSAAVAASPWGMINVNTTNSPRNCLLVNNTSSC
SQVQISSSPRN LGIKRRKSQAKKVCIPAPAAMNSRSSGEVVPSDLWAWRKYGQKPIKGS
PYPRGYRCSSSKSGSARKQVERSRTDPNMLVITYTSEHNHPWPTQORNALAGSTRSSSSS
SLNPSSKSSATAATSPSSRVFQNNSSKDEPNNSNLPSSTHPPFDAAAIKEENVEERQE
KMEFDYNDVENTYRPELLQEFQHPEDFFADLDELEGDSLTMLLSHSSGGGNMENKTIP
DVFSDFDDDESSRSL*

>G1965 (1..609)

ATGGATAACTTCAATGTTGTTGCCAATGAAGACAATCAAGTGAATGATGTGAAGCCTCCA
CCACCCACCACCGAGTGTGTGCAAGATGTGATTCTGATAACACAAAATTTGTTACTAC

AACAATTATAGTGAGTTTCAACCGCGCTACTTCTGCAAGAACTGTCGAAGATACTGGACT
CATGGTGGGGCTTTAAGAAACGTACCAATTGGTGGGAGTAGTCGTGCCAAGCGGACAAGG
ATAAATCAACCTTCAGTTGCTCAGATGGTTTCTGTTGGAATCCAACCAGGGAACCGTTTT
AGTCTTTTGTCTCATATTCATGGTGGTATGGTAACAAATGTGCATCCAACCTCAAACCTTT
CGACCAAATCATCGCCTAGCTTTCCATAATGGATCATTGAGCAAGATTATTATGATGTT
GGGTCTGATAATCTTTTGGTAAACCAACAAGTTGGTGGATATGTTGATAATCACAACGGT
TATCATGAATCAAGTGGATCAATACAACTGGAACCAGAGCTTCAATAACGCTATGAAC
ATGAATTATAATAACGCTAGCACTAGCGGAAGGATGCATCCTAGTCATTTAGAGAAGGGT
GGTCCTTGA

>G1965 Amino Acid Sequence (domain in AA coordinates:27-55)
MDNFNVVANEDNQVNDVKPPPPPPRVCARCDSDNTKFCYNNYSEFQPRYFCKNCRRYWT
HGGALRNVPIGGSSRAKRTRINQPSVAQMVSVGIQGNRFSSLSIHGGMVTNVHPTQTF
RPNHRLAFHNGSFEDYDVGSDNLLVNQQVGGYVDNHNGYHMNQVDQYNWNQSFNNAMN
MNYNNASTSGRMHPSHLEKGGP*

>G1976 (1..1152)

ATGACTGATCCTTATTCCAATTTCTTCACAGACTGGTTCAAGTCTAATCCTTTTCACCAT
TACCCTAATTCCTCCACTAACCCCTCTCCTCATCCTCTTCTCTGTTACTCTCCCTCT
TCCTTCTTCTTCTTCCCTCAATCCGGAGACCTCCGCGCTCCACCGCCGCCACCACTCCT
CCTCCTTCTCCTCCTCCTCGAGAAGCCCTCCTCTCCTCAGCCTCAGCCCCGCCAACAAA
CAACAAGACCACCATCACAAACCATGACCACCTTATTCAAGAACCACCTTCAACCTCCATG
GATGTCGACTACGATCATCACCATCAAGATGATCATATAACCTCGATGACGATGACCAT
GACGTCACCGTTGCTCTTCACATAGGCCTTCCAAGCCCTAGTGCTCAAGAGATGGCCTCT
TTGCTCATGATGTCTTCTTCTCCTCTTCTCCTCGAGGACCACTCATCATCAGGAGCATG
AATCACAAGAAAGACCTCGACCATGAGTACAGCCACGGAGCTGTGCGAGGAGGAGAAGAT
GACGATGAAGATTAGTCCGCGGAGACGGCGGCTGTAGAATCAGCAGACTCAACAAGGGT
CAATATTGGATCCCTACACCTTCTCAGATTCTCATTGGCCCTACTCAGTTCTCATGTCTCT
GTTTGCTTCAAAACCTTCAACAGATAACAATAACATGCAGATGCATATGTGGGGACATGGA
TCACAATACAGAAAAGGACCTGAATCTCTAAGGGGAACACAACCAACAGGAATGCTAAGG
CTTCCGTGCTATTGCTGCGCCCCAGGCTGTGCAACAACATTGACCATCCAAGGGCAAAG
CCTCTCAAAGACTTCAGAACCCCTTCAAACACATTACAAGAGAAAACATGGGATCAAACCT
TTCATGTGTAGGAAATGTGGAAAGGCTTTCGCAGTCCGAGGGGACTGGAGAACACATGAG
AAGAATTGTGGCAAACCTTTGGTATTGCATATGTGGATCTGATTTCAAGCACAAGAGATCT
CTCAAAGATCACATCAAGGCTTTTGGGAATGGTCATGGAGCCTACGGAATTGATGGGTTT
GATGAAGAAGATGAGCCTGCCTCTGAGGTAGAACAATTAGACAATGATCATGAGTCAATG
CAGTCTAAATAG

>G1976 Amino Acid Sequence (domain in AA coordinates: 219-323)
MTDPYSNFFTDWFKSNPFHYPNSSTNPSPHPLPPVTPPSSFFFFPQSGDLRRPPPPPTP
PPSPPLREALPLLSLSPANKQQDHHHNDHLIQEPPSTSMVDVYDHHHQQDHHNLDHDDH
DVTVALHIGLPSPSAQEMASLLMSSSSSSSRTHHHEDMNHKKDLDEYSHGAVGGGED
DDEDSVGGDGGCRISRLNKGQYWIPTPSQILIGPTQFSCPVCFKTFNRYNNMQMHMWHG
SQYRKGPESLRGTQPTGMLRLPCYCCAPGCRNNIDHPRAKPLKDFRTLQTHYKRKHGIXP
FMCRCGKAFVRGDWRTHKNCGLWYCICGSDFKHKSRLKDHIAKAFNGHGHAYGIDGF
DEEDEPASEVEQLDNDHESMQSK*

>G2057 (27..1289)

GCCGTCTCGACGAATATGCTCTACCAATGTCTGACGACCAATTCCATCACCCGCGCCTC
CTTCTTCAATGAGGCACCGTTCTACGTGCGATGCGCGGACGGCGGCTGCGGCGAGATTG
TTGAGGTGCAAGGTGGTCACATTGTTTGGTCTACCGGAAGAAAAGACCGCCACAGCAAAG
TCTGCACGGCTAAAGGGCCACGTGACCGGCGGTGAGACTCTCTGCTCACACGGCGATTCT
AGTTTTACGATGTTCAAGACAGGCTTGGTTTTCGACCGACCTAGCAAAGCCGTTGATTGGC
TTATCAAAAAGGCTAAGACTTCCATTGACGAGCTCGCTGAGCTTCTCCTGGAATCCCG
CCGATGCAATTCGCCCTAGCCGCTGCTAACGCTAAACCCAGAAGAACCACCGCCAAAACCC
AAATCTCTCCGTCTCCGCCACCGCCGCAACAGCAACAACAACAACAGCTTCAGTTTCG
GTGTTGGCTTCAACGGAGGAGGAGCAGAGCATCCGAGTAACAACGAGTCGAGTTTTCTCC
CGCCGTCAATGGATTTCAGATTCTGATGACTATAAAGTCGTTTTTCCGGTGATTG
GCTCTTCAACGGAGGCTCCTTTCGAATCATACCTTATGCACAACATATCATCATCAGCATC
CGCCGGATTGCTTCTCTGAACTAATAGCCAAAACCAAGATCTCCGTCTCTCGCTGCAAT
CGTTCCCGGATGGTCCACCGTCTGCTTCTGCAACCACCAACATCACCACCACCTCTGCTT

CCGCCTCCGAGCCTACTCTGTTCTACGGACAGAGCAATCCGTTAGGGTTTGACACATCGA
 GTTGGGAGCAGCAGTCGTCGGAATTCGGAAGGATTAGAGACTAGTGGCTTGAACAGCG
 GCGGTGGCGGCGGAGCAACCGATACAGGAAACGGAGGAGGGTTTCTGTTCCGCTCCTCCTA
 CTCCTTCAACGACGTCGTTTCAGCCAGTTCTTGGCCAAAGCCAACAGCTTTATTCTCAGA
 GGGGTCCCTTTCAGTCCAGTTACAGTCCCATGATCCGTGCTTGGTTTGATCCTCACCATC
 ATCACCATCCATCTCCACCGACGATCTCAACCACCACCATCACCTTCCTCCACCGGTTTC
 ACCAATCAGCAATCCCCGAATCGGATTCGCCTCAGGTGAATTCTCTTCGGGTTTTTCGCA
 TACCAGCAGGTTTTAGGGCCAAGAAGAGGAGCAGCAGCAGGTTCTACTCACAAGCCGT
 CCTCTGCTTCTCTATTTCTCGCCATTGACAATCGAACTAATCCTC

>G2057 Amino Acid Sequence (domain in AA coordinates: TBD)
 MSDDQFHHPPPPSSMRHRSTSDAADGGCGEIVEVQGGHIVRSTGRKDRHSKVCTAKGPRD
 RRVRLSAHTAIQFYDVQDRLGFDPRPSKAVDWLIKAKTSIDELAEPLPPWNPADAIRLAAA
 NAKPRRTAKTQISPSPPPPQQQQQQQLQFVGFGNGGAEHPSNNESSFLPPSMDSDSI
 ADTIKSFPPVIGSSTEAPSNHNLHMHYHHQHPPDLLSRTNSQNQDLRLSLQSFDPGPPSL
 LHHQHHTSASASEPTLFYQSNPLGFDTSSEWQSSSEFGR.IQRLVAWNSGGGGGATDT
 GNGGGFLFAPPTPSTTSFQPVLGQSQQLYSQRGPLQSSYSPIRRAWFDPHHHHQSISTDD
 LNHHLHPPPVHQAIPGIGFASGEFSSGFRIARFQGEQEEQHDLGLTHKPSSASSISRH
 *

>G2107 (79..624)

ACCACAAAACAGAGCAACACACAACAAAGCTTCATTTCAATTCTGTTTCGAGAACCCT
 TTGAGAACGAGATCGGAGATGGAACACGACGATACACCGTGGCGGAGATGAAGCCAAAG
 AAGCGTCTGGACGGAGGATTTTCAAGGAGACACGTACCCAATCTACAGAGCGTGCAGG
 CGTAGGGACGGCGACAAATGGGTATGCGAAGTCCGTGAACCGATTATCAGCGTCGAGTC
 TGGCTCGGAACCTATCCGACGGCAGATATGGCCGCACGTGCTCACGACGTGGCGGTTCTT
 GCTCTGCGCGGAGATCCGCGTGTGTTGAATTTCTCCGATTCTGCTTGGAGGTTGCCGGTG
 CCGGCATCCACTGATCCGGACACGATCAGGCGCACGGCGGCCGAAGCAGCGGAGATGTTT
 AGGCCCGCGGAGTTTAGTACAGGAATTACGGTTTTACCTCAGCCAGTGAGTTTGACACG
 TCGGATGAAGGAGTCGCTGGAATGATGATGAGGCTCGCGGAGGAGCCGTTGATGTCGCCG
 CCAAGATCGTACATTGATATGAATACGAGTGTGTACGTGGACGAAGAAATGTGTTACGAA
 GATTTGTCACTTTGGAGTTACTAAATACGTATGTGTTAAAAACCAAAGATCGTATGTG
 TATGTATGCATAATAAATGGGCTTAATGATGGGCATAGATATGATAGGTCCAGCCTATAT
 GTTAAATGTGTTTTATTTTTTGGTTTATCTAGTTTCTAGGTATTTACCAAATTGTATTA
 GTATAAGTTTTATTAAGAAATAATCAAAATGTTGTTGCCAAAAAAAAAAAAAAAAAAAA
 AAAAA

>G2107 Amino Acid Sequence (domain in AA coordinates: TBD)
 MENDDITVAEMKPKKRAGRRIFKETRHPIYRGVRRRDGDKWVCEVREPIHQRRVWLGTYP,
 TADMAARAHDAVLALRGRSACLNFSASAWRLPVPASTDPDTIRRTAAEAEMFRPPEFS
 TGITVLPSASEFDTSDEGVAGMMMLRLAEPLMSPPRSYIDMNTSVYVDEEMCYEDLSLWS
 Y*

>G211 (1..750)

ATGATGTCATGTGGTGGGAAGAAGCCAGTGTCTAAGAAAACAACGCCGTGTTGCACGAAG
 ATGGGGATGAAGAGAGGACCATGGACGGTGGAGGAAGACGAGATTCTTGTGAGCTTCATT
 AAGAAAGAAGGTGAAGGACGGTGGCGATCGCTTCTAAGAGAGCTGGTTTACTCAGATGT
 GGAAAGAGCTGTCGTCTACGGTGGATGAATCTCCGACCCCTCGGTTAAACGTGGAGGA
 ATTACGTCCGACGAGGAAGATCTCATCTCCGTCTTCACCGCTCCTCGGCAACAGGTGG
 TCATTGATCGCGGGAAGGATAACCGGAAGGACTGATAATGAAATTAAGAACTATTGGAAC
 ACTCATCTTCGTAAGAACTTTTAAGGCAAGGAATTGATCCTCAAAACCACAAGCCTCTT
 GATGCAAAACAACATCATAAAGCAAGAAAGTTCGGGTGGACAAAAGTACCCCTCTA
 GAGCCTATTTCTAGTTCTCATACTGATGATACCACTGTTAATGGCGGGGATGGAGATAGC
 AAGAACAGTATCAATGCTTTGGTGGTGAACACGGCTACGAAGACTTTGGTTTCTGCTAC
 GACGACAAGTTCTCATCGTTTCTTAATTCGCTCATCAACGATGTTGGTGATCCTTTTGGT
 AATATTATCCCAATATCTCAACCTTTGCAGATGGATGATTGTAAGGATGGGATTGTTGGA
 GCGTCGTCTTCTAGCTTAGGACATGACTAG

>G211 Amino Acid Sequence (conserved domain in AA coordinates: 24-137)
 MMSCGGKKPVSKKTPCCTKMGMKRGPWTVEEDILVSFIKKEGEGRWRSLLPKRAGLLRC
 GKSCRLRWNNYLRPSVKRGITSDEEDLILRLHLLGNRWSLIAGRIPGRTDNEIKNYWN
 THLRKLLRQIDPQTHKPLDANNIHKPEEEVSGGQKYPLEPISSHTDDTTVNGGDGDS

KNSINVFGGEHGYEDFGFCYDDKFSSFLNSLINDVGDPFGNIIPISQPLQMDCKDGIVG
ASSSSSLGHD*

>G2133 (26..457)

ATCTCATCTTCATCCACCCAAAAACATGGATTCAAGAGACACCGGAGAACTGACCAGAG
CAAGTACAAAGGTATCCGTCGTCGGAAATGGGGAATGGGTATCAGAGATTCGTGTCCC
GGGAACCTCGTCAACGTCTCTGGTTAGGCTCTTTCTCCACCGCAGAAGGCGTGCCTAGC
CCACGACGTCGCTTTTTTACTGCTTGACCCGACCATCTCCCTCGACGACGAATCTTTTAA
CTTCCCTCACTTACTTACAACCTCCCTCGCCTCCAATATATCTCCTAAGTCCATCCAAAA
AGCTGCTTCCGACGCCGGCATGGCCGTGGACGCCGATTCCATGGTGTGTGTCTGGGAG
TGGTGGTTGTGAAGAGAGATCTTCCATGGCGAATATGGAGGAGGAGGACAACTTAGTAT
CTCCGTGTATGATTATCTTGAAGACGATCTCGTTTGATCTATACGAGTACGTTTTTAGCA
GTTAA

>G2133 Amino Acid Sequence (domain in AA coordinates:11-83)
MDSRDTGETDQSKYKGIIRRRKWKWVSEIRVPGTRQRLWLGSFSTAEGAAVAHDVAFYCL
HRPSSLDDSFNFPHLTSLASNISPKSIQKAASDAGMAVDAGFHGAVSGSGGCEERS
MANMEEEDKLSISVYDYLEDDL*

>G2134 (36..644)

GAGCAAAAACCTTTGTGTGCGTGTGTGTGTGTGTTTCATGGCTGGTCTTAGGAATTCGGTA
ACAGCGACAAAGCGCAAAACGATGGCAAAGGTGTACCATCTGCCTACAGAGGAGTCCGGA
AGAGAAAATGGGGGAAATGGGTGTCTGAAATCCGTGAACCGGGGACCAAGAACCGTATCT
GGCTAGGCAGTTTCGAGACTCCTGAAATGGCTGCAACCGCATACGACGTGGCAGCATTTCT
ATTTTCAGAGGGAGAGAAGCTCGTCTCAACTTCCCTGAGCTCGCCAGCAGCTTCCACGTC
CTGCAGACTCTAGCTCAGACAGCATTCGCATGGCAGTTTCATGAGGCAACACTCTGCCGCA
CCACCGAAGGAACAGAGTCAGCCATGCAAGTGGACAGCTCAAGCTCCTCCAATGTAGCTC
CAACAATGGTCAGACTCTCGCCAGGAAATTCAGCGATCAACGAGTCAACTTTGGGAT
CTCCTACTACAATGATGCATTCAACATACGACCCTATGGAGTTTGCTAATGATGTGGAGA
TGAATGCTTGGGAAACATACAGAGTGACTTTCTTTGGGACCTTAACCCCAAAACCTAA
CTCATGGAGAGCTTCTACAGCTCAATCTTACAATACCAGCATAAGTTACTGGCTTAGAAT
ACTTAAATTTATTGAAGTTTAGTTTTCAGAGTCTACCACAAGGGTTGTTGATTCTGACGT
TATAGCAAAGAATAAAGCTCATCAGATTTTGGAGGAAAGACTCTATGAGCTTGATGGGT
CCCTGAAAGGACCTCTTCACAAATATTTTAAATTTTGTGTTACTAGTAGAAACATAGA
TTATGAGGTGTGACTTATTATTATTTTACAATTGTTTGTACCTCATTGATGTATTG
ATTT

>G2134 Amino Acid Sequence (domain in AA coordinates: TBD)
MAGLRNSGNSDKAQNKGKVPVPSAYRGVRRKWKWVSEIREPGTKNRIWLGSEFTEPMAA
TAYDVAAAFHFRGREARLNFPELASSLPADSSSDSIRMAVHEATLCRTTEGTESAMQVD
SSSSSNVAPTMVRLSPREIQAINESTLGSPTTMMHSTYDPMFANDVEMNAWETYQSDFL
WDP*PQNLTHGELLQLNLTPA*

>G2151 (236..1321)

TTTTTTTTTTAGGGTTCATAAGAACAATTGGATTTTGAGCTCACAGTATAAATAACCCG
ACTTTGATTACTGGTAATTTTAAACCGCCATTGTTGTTCTCTTACTACTTTTGGGAA
TTAGGGTTTATGATTTCTGGGTATTAGATTAGATAAATTTGTTTCTTTTGTTAATC
AATTTAAAAATCTCTTATTTCTGTTAAAGACTTGTAATTTTGGAGTTTTTAATGCATGGA
CGGAAGAGAAGCAATGGCATTTCAGGCTCGCATTCTCAGTACTATCTTCAAAGAGGAGC
CTTTACTAATCTCGCACCTTCCCAAGTCGCGAGTGGGCTTCACGCGCCGCCACATAC
GGGATTGAGGCCAATGTCTAACCCCTAACATTATCACCCTCAGGCTAACATCCAGGACC
TCCTTTCTCGGATTTTGGACACACCATTCACATGGGAGTGGTCTCCTCTGCTTCTGATGC
TGATGTGCAACCGCCACCAGCCACCAGAGGAAACCGATGGTTAAGAGGAAACG
TGGACGGCCAAGAAAGTATGGAGAACCAGTGGTTAGTAATAAGTCTAGGGACTCTTCTCC
AATGTCGTATCCTAATGAACCTAAACGGGCCAGAGGTCGACCTCCTGGAACCTGGAAGGAA
GCAACGCTTGGCTAATCTTGGTGAGTGGATGAATACTTCAGCTGGACTTGCTTTTGACAC
TCATGTGATCAGCATTGGAGCAGGAGAAGACATTGCTGCGAAAGTTTTGTCAATTTTACA
ACAAAGACCTCGGGCTCTTTGTATAATGTGAGGCACTGGAACATTTCTTCAGTCACTCT
GTGCAAAACCGGTTCAACCGATCGTCACTTAACATACGAGGGACCTTTTGAGATTATAAG
TTTTGGTGGATCTTATTTGGTGAATGAAGAAGGTGGATCCAGAAGTCGAACAGGCGGATT
GAGTGTCTCTCTTCTCGTCCCGATGGTAGTATTATGCGGTTGGAGTTGACATGCTTAT
CGCAGCCAACCTTGTTCAGGTGGTGGCATGTAGTTTTGTATACGAGCAAGGGCAAAGAC

TCATAATAACAATAACAAGACCATCAGACAAGAAAAGGAACCAATGAAGAGGACAACAA
TAGTGAAATGGAGACCACACCGGGTAGTGCAGCTGAACCAGCAGCATCTGCGGGTCAGCA
GACGCCACAGAACTTCTCTTCTCAGGGAATAAGGGGGTGGCCCGGTTTCAGGCTCAGGCTC
TGGCAGATCACTTGCACATTTGCAGAAACCCACTCACTGATTTTGATTGACTCGTGGATG
ATATACACTATTAGTCTTTGAAGCAGCAGCATACAAAATGTGATTGCTGTACATATGTTA
TTGTAGATTCTCTCTGGAATGTTGAAATCAGACATTTAAGGATTGATACTAGATCTCT
CAGCTCCTTCTAACATTGTTAATGTAACAGAACCTCCCACTTTTCATGCTATTTGC

>G2151 Amino Acid Sequence (domain in AA coordinates:93-113, 124-144)

MDGREAMAFPGSHSQYYLQRGAFNLAPSQVAGSLHAPPPHTGLRPMSPNIHHPQANNP
GPPFSDFGHTIHMVSSASDADVQPPPPPPPEEPMVKRKRGRPRKYGPMVSNKSRDS
SPMSDPNEPKRARGRPPTGRKQRLANLGEWMNTSAGLAFAPHVISIGAGEDIAAKVLSF
SQQRPRALCIMSSTGTTISSVTLCKPGSTDRHLTYEGPFEIISFGGSYLVNEEGGSRSTG
GLSVLSLRPDGSI IAGGVDMLIAANLVQVVACSFVYGARAKTHNNNNKTIHQEKEPNEED
NNSEMETTPGSAEPAASAGQOTPQNFSSQGIKRGWPGSGSGSGSLDICRNPLTDFDLTR
G*

>G2154 (82..1317)

GCAAAAAGAAAAAATGAAAAAATCCCTAACTCTCTCTCTCTAGAAATTCTTATTTTGG
TGCGTATCTCTCTAAAAAGGAATGGATCCTAACGAAAGCCACCATCACCACCAACAACAA
CAGCTCCATCACCTCCACCAACAGCAACAGCAACAGCAGCAGCAACGACTCACTTCT
CCTTACTTCCACCACCACTACAGCACCATCACCACCTTCCAACCACCGTAGCAACCACC
GCTTCTACCGGAAACGCGGTTCCATCTTCCAACAATGGGCTTTTCCCTCCGCAGCCTCAG
CCACAGCACCAGCCTAATGATGGGTCACTTCTCTCGCGGTGTACCCTCATTCACTTCCG
TCCTCGGCTGTGACGGCGCGGATGGAGCCGGTAAAGAGGAAGAGGGGTCGACCAAGAAAG
TATGTGACGCCGGAACAAGCCCTAGCGGCTAAGAAATTGGCGTCTTCTGCGAGTAGTTCG
TCTGCTAAACAGAGCGGAGAGCTTGTCTGTGTTACCGGTGGTACGGTATCGACTAATTCC
GGGTCACTCAAGAAATCTCAGCTTGGTCTGTCTCGGGAACCTGGACAATGTTTACTCCG
CATATTGTTAATATAGCTCCTGGCGAGGATGTGGTCCAGAAAATTATGATGTTGCAAC
CAAAGCAAGCATGAATATGCGTCTTCTTCTGCATCAGGCACTATCTCTAATGCATCCTTG
CGCCAACCGGCTCCATCAGGAGGCACTTACCATATGAGGGTCAATACGAGATTCTCTCA
CTATCTGGATCTTATATCCGAACCTGAACAAGGTGGTAAATCCGGCGGCCTTAGCGTTTCT
TTATCTGCTTCAGATGGTCAGATCATCGGTGGAGCGATTGGTAGCCATCTCACAGCTGCT
GGCCCCGTTTCAGGTGATTCTTGGTACGTTTTCAGCTTGATAGAAAGAAGGATGCCGCCGGG
AGTGGTGGGAAAGGGGATGCTTCAAACAGTGAAGTTCGGTTAACTTCTCCTGTAAGCTCT
GGACAGTTGCTTGGCATGGGTTTCCCTCCTGGTATGGAATCTACGGGAAGAAATCCAATG
AGGGGAAACGACGAGCAACATGATCATCATCATCAAGCCGGTTTGGGTGGACCTCAT
CATTTTCATGATGCAAGCGCGCGAGGGGATACACATGACACATTCCAGGCCATCTGAATGG
CGCGGAGGAGCAACAGCGGTCATGATGGCAGAGGCGGTGGCGGGTATGATTGTGTCAGGA
AGGATAGGACATGATCGTCGGAGAATGGAGATTACGAGCAGCAATACCGGATTAGCAG
AGCTTCCAGGAGAAGTGTGTAGAGTTTAGATCCCAAGTAGAGAAACAGAAGCGAGCAAA
GAATCTGAAGTGAAGAGGACTTATTAGACAGAGACTCGTCTGAAGGGTCTTTAATCATA
GAAAGAAGTTGCTGAGTGATTGCTTTTGTCTTCTTCTTGGTACGGTGTATTATATTAAC
TCCACAACCTTTTTTTTATACTTTTCAGTAACGATTCTCCTTCACTTTCAATTTTATTCTT
TTTTTTTATACTCTTTTTCTTTCTTATAATATTTTTTTTGGTTTTTCTTTCGTTTGTTA
CTAAAAAGGAAATGCTCTTTTGTGAAATATATACACTTCGTTTG

>G2154 Amino Acid Sequence (domain in AA coordinates:97-119)

MDPNESHHLHQQQLHHLHQQQQQQQQQLTSPYFHHQLQHHLHLPPTVATTASTGNAV
PSSNNGLFPPQPQPHQPNDSGLAVYPHSPSSAVTAPMEPVKRRGRPRKYVTPEQA
LAAKKLASSASSSSAKQRRELAAVTGGTVSTNSGSSKKSQLGSGVKTGQCFTPHIVNIAP
GEDVVQKIMMFANQSKHELCLVLSASGTISNASLRQPAPSGGNLPYEGQYEILSLSGSYIR
TEQGGKSGGLSVLSASDGQIIGGAIGSHLTAAGPVQVILGTFQLDRKKDAAGSGGKGD
SNSGSRLTSPVSSGQLLGMFPFGMESTGRNPMRGNDQHDHHLHQAGLGGPHHFMQAP
QGIHMTSRPSEWRGGGNSGHDGRGGGGYDLSGRIGHSESENGDYEQQIPD*

>G2157 (306..1238)

TCCTTTGATTTTAACTTTTTCAGTAGCAAGCCAAAAAAGGACAGACAAAGAAGTT
CCTTTTATGATAAAGGTATGATGATAGCAAAACAAATGATACCCCATGTCTTGTGTGTCT
GCTTCATGCAACATGTTGGTTTGGATTGGTTAATCTAAAAGTTTAAAGATAAGGTTTTCG
GATTCTCTTCTGTCTTGTAAATAGTTTCTTGTCTCGGAGAGCCATCAACACCAACTTCAACA

AAAAAACAAGAAAAAGAAAAAGATTCTCTTTCTCGTTTTATTTCCATTAGAGAAGAAAA
 AAAGAATGGCGAATCCTTGGTGGGTAGGGAATGTTGCGATCGGTGGAGTTGAGAGTCCAG
 TGACGTCATCAGCTCCTTCTTTGCACCACAGAAACAGTAACAACAACCAACCCACCGACTA
 TGACTCGTTCGGATCCAAGATTGGACCATGACTTCACCACCAACAACAGTGGAAGCCCTA
 ATACCCAGACTCAGAGCCAAGAAGAACAAGACAGCAGAGACGAGCAACAGCTGTTGAAC
 CCGGATCCGGATCCGGGTCTACGGGTCGTCGTCCTAGAGGTAGACCTCCTGGTTCCAAGA
 ACAAAACCAAGAGTCCAGTTGTTGTTACCAAGAAAGCCCTAACTCTCTCCAGAGCCATG
 TTCTTGAGATTGCTACGGGAGCTGACGTGGCGGAAAGCTTAAACGCCCTTTGCTCGTAGAC
 GCGGCCGGGGCGTTTCGGTGCTGAGCGGTAGTGGTTTGGTTACTAATGTTACTCTGCGTC
 AGCCTGCTGCATCCGGTGGAGTTGTTAGTTTACGTGGTCAGTTTGAGATCTTGTCTATGT
 GTGGGGCTTTTCTTCTTACGTCTGGCTCTCCTGCTGCAGCCGCTGGTTTAACCATTTACT
 TAGCTGGAGCTCAAGGTCAAGTTGTGGGAGGTGGAGTTGCTGGCCCGCTTATTGCCTCTG
 GACCCGTTATTGTGATAGCTGCTACGTTTTCGAATGCCACTTATGAGAGGTTACCGATTG
 AGGAAGAACAACAGCAAGAGCAGCCGCTTCACTAGAAGATGGGAAGAAGCAGAAAGAAG
 AGAATGATGATAACGAGAGTGGGAATAACGAAACGAAGATCGATGCAGCCGCCGATGT
 ATAATATGCCTCCTTAATTTTATCCCAAATGGTCATCAAATGGCTCAACACGACGTGTATT
 GGGGTGGTCTCCTCCGCTCGTCTCCTCCTCGTATTGATTAGTTAGATAGGCGGTGGTTG
 GTGCGTTCTTTTACTGGAATGATTATATTTTCCATTAGGATGGTTAGGCTTTTGTATTAT
 TAAAGCTATCAAGTTTCTTTTCTTTTACGGATAATTCGGATGACAATTAGCTAGTGTTT
 GTTGTGTTGTTTGTGGCGGCTTTTCTGACTTGACTATTTTGTATCGCGGATAGCTTTGTA
 TGAAAGTGAATTGATTGTAGAATCGTCTTTTGAATTTTGATGTTGAAAAAACCAA
 >G2157 Amino Acid Sequence (domain in AA coordinates: 82-102, 164-107)
 MANPWVGNVAIGGVESPTSSAPSLHHRNSNNNPPTMTRSDPRLDHDFTTNNSGSPNT
 QTQSQEEQNSRDEQPAVEPGSGSGSTGRRPRGRPPGSKNPKSPVVTKEPNSLQSHVL
 EIATGADVAESLNAFARRRGRGVSVLSGSLVTNVTLRQPAASGVVSLRGQFEILSMCG
 AFLPTSGSPAAAAGLTIYLAGAQGVVGGVAGPLIASGPVIVIAATFCNATYERLPIEE
 EQQQEQPLQLEDGKKQKEENDDNESGNNNGNEGSMQPPMYNMPNFI PNGHQMAQHDVYWG
 GPPPRAPPSY*
 >G2181 (1..1005)
 ATGATGCTTGCGGTGGAAGATGTGTTAAGCGAACTCGCCGGAGAAGAAAGGAACGAGAGA
 GGATTGCCACCTGGCTTCCGGTTTCACCCGACGGACGAAGAGCTCATTACCTTCTACTTA
 GCTTCCAAAATCTTCCATGGTGGTCTCTCCGGCATTACATTTCGAAGTTGATCTCAAC
 CCTGTGGAACCTTGGGAGCTTACCAGAAATGGCGAAGATGGGAGAGAGAGTGGTACTTT
 TATAGTCTAAGGGACAGGAAATATCCGACAGGTTTGAGGACTAACAGAGCAACTACTGCT
 GGATACTGGAAGCTACCGGCAAGATAAGGAAGTCTTCTCCGGCGGAGGAGGACAGCTT
 GTTGGGATGAAGAAGACGTTGGTGTCTACAAAGGTAGGGCTCCACGTGGCCTCAAGACT
 AAGTGGGTGATGCATGATGATATCGCCTCGAAAACGACCATTACACCGCCACACGTGTAAG
 GAGGAATGGGTGATTTCAGAGTGTTCAATAAAACAGGAGACAGAAAAAATGTTGGATTA
 ATCCATAACCAAATCAGCTACCTTCATAACCATTCACTCTCAACAACACATCATCATCAT
 CATGAAGCCTTACCTTTGCTTATAGAACCCTTCAACAAAACCCTAACCAACTTCCCATCA
 CTACTCTACGATGATCCACACCAAACTACAATAATAACAACCTCCTTCATGGATCATCA
 GGCCACAACATCGACGAGCTCAAAGCCTTAATCAACCCTGTCTGCTCTCAGCTCAACGGT
 ATCATCTTTCTTTCAGGGAACAACAACACGACGAAGACGACTTCGACTTTAACCTCGGC
 GTGAAAACAGAGCAGTCTTCAACGGTAACGAAATTGACGTACGAGATTACTTGGAGAAC
 CCTCTGTTTCAGGAAGCGAGTTATGGTCTGTTGGGTTTTTCTGCTTCTCCTGGACCTCTT
 CACATGCTACTAGATTCTCCATGTCTTTTAGGATTCCAGCTGTAG
 >G2181 Amino Acid Sequence (conserved domain in AA coordinates: 22-169)
 MMLAVEDVLSELAGEERNERGLPPGFRFHTDEELITFYLASKIFHGGLSGIHISEVDLN
 RCEPWELPEMAKMGEREWYFYSRLDRKYPTGLRTNRRATTAGYWKATGKDKEVFGGGGQL
 VGMKKTLLVFYKGRAPRGLKTKWVMHEYRLNENHSHRHTCKEWWICRVFNKTGDRKNVGL
 IHNQISYLNHNSLSTTHHHHHEALPLLEPSNKTLTNFPSSLYDDPHQNYNNNNFLHGSS
 GHNIDELKALINPVVSQNLNGIIFPSGNNNNDEDDDFNGLGVKTEQSSNGNEIDVRDYLEN
 PLFQEEASYGLLGFSSSPGPLHMLLDSPCLPLGQL*
 >G221 (115..795)
 CTCTCTTATTCTCTCACTCTTTTTTTTTTATATTCTCTCTCTCTAAATCTATAAAATAT
 ATTTAAAAAATTGATCGTATATAATAAAGTAAATAAAGAATAATAACAAAAAAATGGAG
 AAAAGAGGAGGAGGAAGTAGTGGAGGTTCCGGATCATCAGCAGAAGCAGAAGTGAGAAAA

GGACCATGGACGATGGAAGAAGATCTTATTCTTATCAACTATATCGCCAACACGGCGAT
GGTGTTTGGAATTCTCTCGCCAAATCTGCAGGTCTAAAACGAACCGGGAAAAGTTGCCGG
CTCCGGTGGCTGAAC'TATCTCCGCCCCGACGTACGACGGGGAAACATCACTCCAGAAGAG
CAACTTATCATCATGGAAC'TTCATGCTAAGTGGGGAAAACAGGTGGTGGAAAATCGCCAAA
CATCTTCCAGGAAGAACGGACAACGAGATCAAAAATTTCTGTAGGACAAGAATTCAAAAA
TACATCAAGCAATCGGATGTAACAACAACATCGTCCGTTGGATCTCATCATAGCTCAGAG
ATCAACGATCAAGCTGCAAGCACGTGAGCCATAATGTCTTTTGTACACAAGATCAAGCG
ATGGAGACTTATTCTCTACACCGACATCATATCAACATACCAATATGGAATTCAACTAT
GGTAAC'TATTCGGCCCGGGCAGTGACGGCAACCGTGGATTATCCAGTACCGATGACCGTT
GATGATCAAACCGGTGAAAAC'TATTGGGGCATGGATGATATTTGGTCATCAATGCATTTA
TTGAATGGTAATTGAT'TGATCGGTGGACAAAACATGGAATATTAATTGAGTATTATATAT
GATTTT'TAGGAGTACTATTATTAGTACGTGACATGTATATGTTTTGCCTCGTTGTAGAG
GTTTGGGGTTATAATTAATATATAATGTTATCTAATATGCAACCTTGATACATATTGGA
TCTTTATTGAACCCATGTTATACATAAAATAAAATGTTGAAGGGGTCATAAAAAAAAAA
AAAAAAAAAAAAA

>G221 Amino Acid Sequence (domain in AA coordinates: 21-125)
MEKRGSSSSGSSGSSAEVRKGPWTMEEDLILINIANHGDGVWNSLAKSAGLKRTGKS
CRLRWLNLRPDVRRGNITPEEQLIIMELHAKWGNRWSKIAKHLPGRTDNEIKNFCRTRI
QKYIKQSDVTTTSSVGSHTSSEINDQAASTSSHNVFCTQDQAMETYSPTPTSYQHTNMEF
NYGNYSAAAVTATVDYPVPMVTDDQTGENYWGMDIWSMHLNNGN*

>G2290 (119..982)

TTCTTTCTTTCTTTCTTTCTTTCTTTCAATCAAGAACAACCCCTAGCTCCTCTCTTTTCTC
TCTCTACCTCTCTTTCTCTATCTTCTCTTATCACTACTTCTCTCGCCGATCAATCATCAT
GAACGATCCTGATAATCCCGATCTGAGCAACGACGACTCTGCTTGGAGAGAACTCACACT
CACAGCTCAAGATTCTGACTTCTTTCGACCGAGACACTTCCAATATCCTCTCTGACTTCGG
TTGGAACCTCCACCACTCCTCCGATCATCTCACAGTCTCAGATTGACTCCGATTTAAC
ACAAACCAACCGGAGTCAAACCTACCACCGTCACTTCTTCTTGTTCCTCATCCGCCGCCGT
TTCCGTTGCCGT'TACCTCTACTAATAATCCCTCAGCTACCTCAAGTTCAAGTGAAGA
TCCGCCGAGAACTCAACCGCCTCCGCCGAGAAAACACCACCACCGGAGACACCAAGTGA
GGAGAAGAAGAGCTCAAAGCGAATTGCGCAACCAAGATTGCGATTTCATGACCAAGAG
TGATGTGGATAATCTTGAAGATGGATATCGATGGCGTAAATATGGACAAAAAGCCGTCAA
GAATAGCCCATTCCTCAAGGAGTACTATAGATGCACAAAACAGCAGATGCACGGTGAAGAA
GAGAGTAGAACGTTTCATCAGATGATCCATCGATAGTGATCACAACATACGAAGGACAACA
TTGCCATCAAACATTGGATTCCCTCGTGGTGGAACTCTCACTGCACACGACCCACATAG
CTTCACCTCTCATCATCATCTCCCTCCTCCATTACCAAATCCTTATTATTACCAAGAACT
CCTTCATCAACTTTCACAGAGACAATAATGCTCCTTACCGCGGTTACCCCGACCTACTAC
TGAAGATACACCTGCCGTGTCTACTCCATCAGAGGAAGGCTTACTTGGTGATATTGTACC
TCAAATATGCGCAACCCCTTGAGGTAAGCTTGGTACGTAGCAATAGCTAAGGAGGTGCTA
ACTCATTATATATAGAAGATATTGCAGACCAGAATATGCGCAGGGAGGGTATAACAATAT
GGCGTTGTAACAATGGATCTATATATTACCTCATTGTTGATCAATAGCACACCACCGGTA
CGTTTGCAATTTCTTCATGTATATTTCTTGTATATATGTAGTTATATATCCAGGTATAA
TTTTGATGTAACACAACATTAATCTTAATCGTGATCCATCCACATTTGATGCATGTAT
GTGCACTTAAGAAAAAGAACATGGAGGAAATAACGTTATTTTTTATTATTCT

>G2290 Amino Acid Sequence (conserved domain in AA coordinates:147-205)

MNDPDNPDLSNDDSAWRELTLTAQDSDFDRDTSNLSDFGWNLHSSDHPHSLRFDSDL
TQTTGVKPTT'TVSSSSSAVSVAVTSTNNNP'SATSSSSEDP'AEINSTASAEKTPPPETPV
KEKKKAQKRIRQPRFAFM'TKSDVDNLEDGYRWRKYGQKAVKNSPFFRSYRCTNSR'CTVK
KRVERSSDDPSIVITTYEGQHCHQTIGFPRGGILT'AHDPHSFTSHHHLPPPLPNP'YYYQE
LLHQLHRDNNAPSRLRPTT'EDTPAVSTPSEGLLDIVPQ'TMRNP*

>G2299 (231..941)

GCCAAAATTTTACCAACATTTTCTCTCTCATATCAAAGTTTCTCTCTCATTTCTTCAT
CACACTTCACTGCCCTGTTTTTTTCTCATT'TTGAATAGTTCTCAAAC'TTATATATTTT
TCCCCCTGAAGCCTAGCTATTTCTTTTATTTGCATTAATCTCGGGATCCGAATCGAAAA
AAGCAATCAGAATAATAGACTTGTACGATACTTGTGCC'TAAGCTAACACAATGGCAGAGG
AATACTACAGCCTCCGCTCGGAGAGAGTAACTCAGCTTCTTGTCCCTAACTCGGAGTCTG
ACTCAGTGAGTGACAAAAGCAAAGCTGAGCAAAGCGAGAGAAGACTAAACGTGGGAGAG
ACTCCGGTAAACACCCTGTTTATCGCGGAGTAAGGATGAGGAAC'TGGGAAAATGGGTGT

CGGAGATTCTGTGAGCCGAGGAAGAAATCACGTATTTGGCTGGGAACCTTTCCCGACGCCGG
AGATGGCGGCGCGTGCACACGACGTGGCGGCTCTGAGCATTAAGGAACGGCCGCTATAC
TAAACTTCCCTGAACCTCGCTGACTCATTCCCTCGACCCGTTTCATTAAGCCCTCGAGACA
TTCAGACAGCAGCTCTTAAAGCAGCTCACATGGAACCGACGACGTCTGTTGTTGATGG
CGTCTTCGTCTCGTCTTTGTCTTCTACGTCTTCGCTCGAGTCTCTTGTGTTGGTGTATGG
ACCTCTCGAGGACTGAGTCGGAGGAGCTCGGTGAGATTGTGGAGCTTCCAAGTCTCGGGG
CGAGTTACGACGTGACTCGGCTAACCTTGGGAACGAGTTTGTCTTCTATGACTCAGTTG
ACTACTGTTTATATCCGCCCGCTGGGGACAGTCGTCCGAAGATAACTATGGTCACGGAA
TTAGCCCTAATTTTGCCCATGGCTTGTCTATGGGATCTCTAACAGTTTATTTTGTATCATT
ACCATAATGTTTTGTATAAACAGTTTATTTTGTATCATTGCCATAATGTTTGTGTTAAT
CACGTTTTTAAACCCCTTTGCTGTTTTTGTGTTTTTTTTTGTAGTTTTT
>G2299 Amino Acid Sequence (conserved domain in AA coordinates:48-115)
MAEEYYSRLRSERVQLVLPNSES DSVSDKSKAEQSEKTKRGRDSGKHPVYRGVVRMRNWG
KWFSEIREPRKKSRIWLGTFTPEMAARAHDAALSIKGTAAILNFPPELADSFPRPVLS
PRDIQTAALKAAHMEPTTSFSSSTSSSSSLSSSTSSLESILVMDLSRTESEELGEIVELP
SLGASYDVDSANLGNFVFYDSVDYCLYPPPWGQSSSEDNYGHGISPNFGHGLSWDL*
>G2340 (274..1275)
ATACAAAACCTCCCTCTTCTCTATCTTCTTCATCTTAAAGAAAAAATAAGAGATATTCGTA
AAGAGAGAACACAAAATTTTCAGTTTACGAAAAGCTAGCAAAGTCGAGTATCGAGGAATAA
CAGAATAAGACGTATCTATCCTTGCCCTTAATGTTCTTACCAAAGATCTAGTCCTTTCTT
TGTATGATCGATCCATCACAAAGCCCAACAACAACAACTACATCTCTTCTCTATCTCT
AGCTTCTATTTTTTAATACATTCAAGAATCAAGAATGGTACGGACGCCGTGTTGTAGAGCA
GAAGGGTTGAAGAAAGGAGCATGGACTCAAGAAGAAGACCAAAGCTTATCGCCTATGTT
CAACGACATGGTGAAGGCGGTTGGCGAACCCCTCCGGACAAAGCTGGACTCAAAGATGT
GGCAAAGCTGCAGATTGAGATGGGCGAATTACTTAAGACCTGACATTAAACGTGGAGAG
TTTAGCCAAGACGAGGAAGATTCCATCATCAACCTCCACGCCATTTCATGGCAACAAATGG
TCGGCCATAGCTCGTAAATACCAAGAAGAACAGACAATGAGATCAAGAACCATTGGAAC
ACTCACATCAAGAAATGTCTGGTCAAGAAAGGTATTGATCCGTTGACCCACAAATCCCTT
CTCGATGGAGCCGGTAAATCATCTGACCATTCCGCGCATCCCGAGAAAAGCAGCGTTTCAT
GACGACAAAGATGATCAGAAATCAATAACAAAAGTTGTGAGGATCATCATCAGCTCGG
TTTTTGAACAGAGTAGCAAACAGATTCCGGTCATAGAATCAACCACAAATGTTCTGTCTGAT
ATTATTGGAAGTAATGGCCTACTTACTAGTCACACTACTCCAACCTACAAGTGTTCAGAA
GGTGAGAGGTCAACGAGTTCCTCTCCACACATACCTCTTCGAATCTCCCATCAACCGT
AGCATAACCGTTGATGCAACATCTCTATCCTCATCCAGTTCCTGACTCCCCGACCCG
TGTTTATACGAGGAAATAGTCGGTGACATTGAAGATATGACGAGATTTTCATCAAGATGT
TTGAGTCATGTTTATCTCATGAAGATTTATTGATGTCCGTTGAGTCTTGTGTTGGAGAAT
ACTTCATTTCATGAGGGAATTAATGATCTTTCAAGAGGATAAAATCGAGACGACGTCG
TTTAATGATAGCTACGTGACGCGCATCAATGAAGTTGATGACTCCTGTGAAGGGATTGAC
AATTATTTTGGATGAGTTATATTGATGATGATGAAAATTTGCATTGGCATGTAAATCAA
TTAGAGTTTGATTTGCTATGGTGTGTTTTAGTTTGTGTGTGTAGTGTGTTTCGACCGTCAA
AAAAAAAAAAAAAAAAAAAAAAAAA
>G2340 Amino Acid Sequence (domain in AA coordinates:14-120)
MVRTPCRAEGLKKGAWTQEDQKLIAYVQRHGEQWRTLPDKAGLKRCGKSCRLRWANY
LRPDIKRGEFSQDEEDSIINLHAIHGNKWSAIARKIPRRTDNEIKNHNWTHIKKCLVKKG
IDPLTHKSLLDGAGKSSDHSAPKSSVHDDKDDQNSNNKLSGSSSARFLNVRVANRFGH
RINHNVLSDIIGSNGLLTSHTPTTSVSEGERSTSSSSTHTSSNLPINRSITVDATSLSS
STFSDSPDPCLYEEIVGDIEDMTRFSSRCLSHVLSHEDLLMSVESCLENTSFMREITMIF
QEDKIETTSFNDSYVTPINEVDDSCGIDNYFG*
>G2346 (1..1011)
ATGGAGTTGTTAATGTGTTCCGGGTCAGGCCGAGTCAGGTGGTTCTTCTTCCACCGAGTCT
TCTTCACTCAGTGGTGGACTCAGGTTTGGTCAGAAGATCTACTTCGAGGATGGATCCGGA
TCCAGAAGCAAGAACCGGTCATACCGTTCCGTAAGTCGTCTACCACGGCGAGGTGCCAA
GTGGAAGGTGTGATAGATGATCTAAGCAATGTTAAAGCTTATTACTCGAGACACAAAGTT
TGTGTCATTCACCTCTAAATCATCTAAAGTCATTGTCTCTGCTCTTCAAAAGGTTTGT
CAACAATGTAGCAGGTTTACCAGCTTTCTGAGTTTGAAGTGGAGAAAAGAAGTTGTGCG
AGAAGACTCGCTTGTCTAACAACGACGAAGAAAACCAACCCCAACACGGCTCTTTTC
ACTTCTCATTACTCTCGAATCGCTCCATCTCTTTACGGAACCCCAATGCTGCAATGATT

AAAAGCGTTTTGGGAGATCCTACTGCGTGGTCAACCGCAAGATCAGTGATGCAGCGGCCT
GGACCGTGGCAGATTAATCCAGTTAGGAAAACCCATCCACACATGAATGTTTTATCACAT
GGAAGCTCAAGCTTTACTACATGTCCAGAGATGATAAACACAATAGCACAGATTCAAGC
TGTGCTCTCTCTCTCTGTCAAACCTACATCCCAATTCATCAGCAGCAACTTCAGACACCA
ACAAATACATGGCGACCATCTTCTGGTTTCGACTCGATGATCTCATTCTCCGATAAGGTT
ACAATGGCTCAGCCACCGCCCATTTCAACCCATCAGCCGCCCATCTCAACACATCAGCAG
TACCTCAGCCAACTTGGGAAGTCATCGCGGGCGAAAAGAGCAATTCACATTATATGTCT
CCTGTGAGTCAAATCTCGGAGCCAGCAGATTTCCAGATAAGCAATGGCAGTGTGTGCGCC
TATTCTCCTCCGTCCTTACTATCTCTTGTGTGCTACTTGC GGCCGCTATAG

>G2346 Amino Acid Sequence (domain in AA coordinates: 59-135)
MELLMCSGQAESGSSSTESSLSGGLRFGQKIYFEDGSGSRSKNRVNTVRKSSITTARCO
VEGCRMDLSNVKAYYSRHKVCCIHSKSSKIVISGLHQRFQCCSRFHLSEFDLEKRSCR
RRLACHNERRRRKPQPTTALFTSHYSRIAPSLYGNPNAAMIKSVLGDPTAWSTARSVMQRP
GPWQINPVRETHPHMNVLSHGSSSFTTCEMINNNSTDSSCALSLLSNSYPIHQQLQTP
TNTWRPSSGFDMSISFSDKVTMAQPPPISTHQPPISTHQYLSQTWEVIAGEKSN SHYMS
PVSQISEPADFQISNGSVSPSPSLLSLVCYLRL*

>G237 (1..852)
ATGGCGAAGACGAAATATGGAGAGAGACATAGGAAAGGGTTATGGTCACCTGAAGAAGAC
GAGAAGCTAAGGAGCTTCATCCTCTCTTATGGCCATTCTTGCTGGACCACTGTTCCCATC
AAAGCTGGGTTACAAAGGAATGGGAAGAGCTGCAGATTAAGATGGATTAAATTACCTAAGA
CCAGGGTTAAAGAGGGATATGATTAGTGCAGAAGAAGAAGAGACTATCTTGACGTTTCAT
TCTCCCTTGGGTAAACAAGTGGTCGCAATAGCTAAATTCTTACCGGAAGAACAGACAAT
GAGATAAAGAACTATTGGCACTCTCATTTGAAAAAGAAATGGCTCAAGTCTCAGAGCTTA
CAAGATGCAAAATCTATTTCCCTCTCTCGTCTTCATCATCATCACTTGTTGCTTGTGGA
GAAAGAAATCCGGAACCTTGATCTCGAATCACGTGTTCTCCCTCCAGAGACTTCTAGAG
AACAAATCTTCATCTCCCTCACAGAAAGCAACGGAAATAACAGCCATCAATGTTCTTCT
GCTCCTGAGATTCCAAGGCTTTTCTCTCTGAATGGCTTTCTTCTCATATCCCCACACC
GATTATTCCTCTGAGTTTACCGACTCTAAGCACAGTCAAGCTCCAAATGTCGAAGAGACT
CTCTCAGCTTATGAAGAAATGGGTGATGTTGATCAGTTCATTACAACGAAATGATGATC
AACACAGCAACTGGACTCTTAACGACATTGTGTTTGGTTCCAAATGTAAGAAGCAGGAG
CATCATATTTATAGAGAGGCTTCAGATTGTAATTCTTCTGCTGAATTCTTTCTCCACCA
ACAACGACGTAAATTGCGTTTATTGTAATGTAAATCAAATTTCTAAGGCAAAACCGGAAA
AAAAAAAAAAAAAAAAAAAA

>G237 Amino Acid Sequence (domain in AA coordinates: 11-113)
MAKTKYGERHRKRLWSPEDEKLRSFILSYGHSCWTTVPKAGLQRNGKSCRLRWINYLR
PGLKRDMSIAEEEEITLTFHSP LGNKWSQIAKFLPGRTDNEIKNYWHS LKKKWLKSQSL
QDAKSISPPSSSSSSSLVACGERNPETLISNHVFS LQRLLENKSSSPSQESNGNNSHQCSS
APEIPLRFSEWLSSSYPHTDYSSEFTDSKHSQAPNVEETLSAYEEMGDVDQFHYNEMMI
NNSNWLNDIVFGSKCKKQEHHIYREASDCNSSAEFFSPPTTT*

>G2373 (48..1199)
GCAAAATCCTCAGATCGTCTTACCTTCTCCGAATCGATCGATTTTTCATGGAGGACGACG
ACGAGATTCAGTCAATTCATCTCCGGGAGATTCTTCCCTTTTACCACAAGCTCCTCCTT
CTCCGCCGATTTTGCCAAACAAACGACGTGACGGTGGCCGTCGTGAAGAAACCACAACCGG
GGCTTTCTTCTCAATCTCCGTCCATGAACGCTTTAGCGTTAGTGGTTCATACTCCTTCTG
TAACCGGTGGTGGTGGTAGCGGAAACAGAAACGGACGAGGAGGAGGAGGAAGCGGTG
GTGGTGGAGGAGGAAGAGATGATTGTTGGAGCGAAGAAGCTACAAAGGTTCTAATCGAAG
CTTGGGGAGATCGATTCTCTGAACAGGTAAAGGAACCTTTGAAGCAACAACATTGGAAAG
AAGTAGCTGAGATTGTGAACAAGAGTCGTCAATGCAAAATACCCTAAAAC TGATATTCAGT
GTAAGAACAGAATTGATACGGTGAAGAAGAAGTATAAGCAAGAGAAAGCTAAGATTGCTT
CTGGTGATGGACCTAGTAAATGGGTTTTCTTCAAGAAGCTTGAGAGTTTGATTGGTGGTA
CTACAACATTTCATTGCTTCTTCAAAAGCTTCAGAGAAGGCTCCTATGGGAGGAGCTCTTG
GGAATAGCCGTTTCGAGTATGTTTAAACGGCAAACCTAAAGGTAATCAGATTGTGCAGCAAC
AACAAAGAGAAGAGAGGCTCTGATTCGATGCGGTGGCATTTTAGGAAACG TAGTGCTTCTG
AGACTGAGTCTGAGTCTGATCCTGAACCTGAGGCTTCTCCTGAGGAATCTGCTGAGAGTC
TCCACCTTTGCAACCGATTCAACCGCTTTCGTTTCATATGCCAAAGCGGTTGAAGGTGG
ATAAGAGTGGAGGTGGAGGAGTGGAGTGGAGATGTGGCGAGGGCGATACTTGGATTTA
CGGAAGCTTATGAGAAGCGGAAACTGCTAAGCTTAAGTTAATGGCGGAAC TGGAAGG

AGAGGATGAAATTTGCTAAAGAGATGGAGTTGCAGAGAATGCAGTTCTTGAAAACCTCAAT
TGGAGATAACACAGAACAAATCAAGAAGAGGAAGAGAGGAGCAGGCAGCGAGGAGAAAGGA
GGATCGTTGATGATGATGATGATCGCAATGGCAAGAATAACGGCAATGTAAGTAGCTGAC
AATTGAACACACAAATGTTCTATGATATTTGCTATGATAAGCTGGATTTTAGGTTTTGA
TGG

>G2373 Amino Acid Sequence (domain in AA coordinates:290-350)
MEDDDDEIQSIPSPGDSSLSPPQAPPSPPIPTNDVTVAVVKPQPGLSSQSPSMNALALVV
HTPSVTGGGSGNRNGRGGGGSGGGGGRDDCWSEETKVLIEAWGDRFSEPGKGTLLKQ
QHWKEVAEIVNKSQCKYPKTDIQCKNRIDTVKKKYKQEKAKIASGDGPSKWVFFKKLES
LIGGTTTTFIASSKASEKAPMGGALGNSRSSMFKRQTKGNQIVQQQEKRGSDSMRWHFRK
RSASETESESDPEPEASPEESAESLPPLQPIQPLSFHMPKRLKVDKSGGGSGVGDVARA
ILGFTEAYEKAETAKLKLMAELEKERMKFAKEMELQRMQFLKTQLEITQNNQEEERSRQ
RGERRIVDDDDDRNGKNNGNVSS*

>G2376 (39..1370)

CACGAGCTTCGACTCAGATCCGGCGATATCGAATTCATGGAGGACGATGAAGACATCC
GATCTCAGGGTTCGATTACCTGATCCGTCTTCTCCCCCGCGGGGACGAATCACGG
TTACGGTGGCTTCGGCAGGTCCGCCCTTCTTATTCTCTGACTCCTCCGGGTAATTCGTCCG
AGAAGGATCCGGATGCGTTGGCTCTGGCGCTGCTTCCGATTACAGCCAGCGGTGGAGGGA
ATAACAGCAGTGGGAGACCAACCGGCGCGCGGAGGAGGATTGTTGGAGCGAAGCAG
CTACGGCTGTGTTGATTGATGCGTGGGGTGAGAGATACTTGGAGCTTAGCAGAGGGAATC
TGAAGCAGAAGCACTGGAAAGAGGTGGCTGAGATTGTGAGCAGCAGAGAGGATTACGGTA
AAATTCCTCCAAACTGATATACAGTGTAAGAATAGGATCGATACGGTGAAGAAGAAGTATA
ACAAGAGAAGGTGAGAATCGCTAACGGCGGTGGCCGTAGCAGATGGGTGTTCTTCGACA
AGCTTGACCGTCTGATTGGATCAACGGCGAAGATCCCGACGGCAACTTCTGGAGTCAGCG
GTCTGTGTCGGAGGATTGCATAAGATTCTATGGGTATTCCAATGGGAAGTCGTTTCAATC
TGTACCATCAGCAAGCTAAGGCTGCAACACCGCCTTTCAATAATCTTGACCGGTTAATTG
GAGCTACGGCTAGAGTCTCAGCTGCTTCTTTCGGTGGCAGTGGTGGAGGAGGCGGAGGAG
GATCTGTCAATGTACCTATGGGAATTCGATGAGTAGCCGTTTCACTCCGTTTGGACAGC
AAGGGAGGACTCTGCCACAGCAAGGTAGGACACTGCCACAGCAACAGCAGCAAGGGATGA
TGGTGAAGAGGTGTAGTGAGTCGAAACGCTGGCGTTTCAGGAAGAGGAACGCTTCTGATT
CAGACTCGGAATCTGAAGCAGCAATGTGATGATTCCGGTGACAGTTTACCACCTCCTC
CTCTGTGCGAAGAGGATGAAGACGGAGGAGAAGAAGAAGCAAGATGGTGTATGGAGTGGGGA
ACAAATGGAGGGAGCTGACTCGGGCAATCATGAGATTCCGTGAAGCTTATGAGCAAACAG
AGAATGGGAACTGCAACAGGTGGTGGAGATGGAGAAAGAGAGGATGAAGTTCTTGAAGG
AGCTTGAGTTGCAGAGAAATGCAGTTCTTTGTGAAGACTCAATTGGAGATATCACTAATTA
AGCAGCAACATGGGAGGAGAAATGGGAAACACCAAGTAATGATCATCATCACAGCCGCAAGA
ACAACATCAATGCGATTGTCAACAACAACAGATTGGGTAATAACTAGAATTTAGTGA
TGCAGTGTGTAATTGATATATTTAGATTTGAG

>G2376 Amino Acid Sequence (domain in AA coordinates:79-178, 336-408)

MEDDEDIRSQSDSPDPSSPPAGRITVTVASAGPPSYSLTPPGNSSQKDPDALALALLP
IQASGGGNSSSRPTGGGGREDWCSEAAATAVLIDAWGERYLELSRGNLQKQHWKEVAEIV
SSREDYGKIPKTDIQCKNRIDTVKKKYKQEKVRIANGGGRSRWVFFDKLDRLLIGSTAKIP
TATSGVSGPVGGLHKIPMGIPMGSRSNLYHQAKAATPPFNNLDRLLIGATARVSAASFGG
SGGGGGGGSVNPVPMGIPMSSRSAPFGQQGRTLPPQQGRTLPPQQQGMVVKRCSESKRWRF
RKRNASDSDSESEAAAMSDSGDSLPPPPPLSKRMKTEKKKQDGDGVGNKWRELTRAIMRF
GEAYEQTENAKLQVVEKERMKFLKELELQRMQFFVKTQLEISQLKQHGRRMGNTSN
DHHHSRKNINAIIVNNNDLGN*

>G24 (194..724)

CGGACGCGTGGGCAAATATTAAAATAAAAAGTGTGGTGAATTCTCAATCTTTGTCTTCT
TTCGTGCTCTCTTTAAACTCCTCCGTCCCTCCTTATTATGTAACCGTCTCGCCGTCAA
TTTCAAAATCTCTCCCTCCGTTCATAAACCCAGATCGAAATTTATGGTTTTGTAATTTT
TTTACCGCGGTTTATGGAGACGGAAGCGGCGGTGACAGCGACGGTTACGGCGGCGACGAT
GGGGATTGGGACGAGGAAGAGAGATCTGAAACCGTATAAAGGAATACGAATGAGGAAATG
GGGGAATGGGTGGCGGAGATACGGGAACCGAATAAGAGATCAAGGATCTGGTTAGGTTT
TTATGCGACGCTGAAGCGGCGGCGAGAGCTTACGACACTGCTGTTTTTTTACCTCCGTGG
TCCTTCAGCGAGGCTTAATTTTCCGGAGCTTTTGGCTGGACTTACTGTTTTTCAACGGCGG
AGGAAGAGGTGGTGAATTTATCGGCGGCGTATATTAGGAGAAAAGCGGCGGAGGTGGTGC

TCAGGTTGATGCGCTTGGAGCGACGGTGGTTGTGAATACCGGCGGCGAGAATCGCGGTGA
TTACGAGAAGATTGAGAATTGTCGTAAGAGCGGTAACGGGTCATTGGAACGGGTCGATTT
GAATAAATTACCCGACCCGAAAATTCCGATGGTGATGATGACGAATGTGTGAAAAGAAG
ATAGAAAAAATAAAAAGTAGCTTGTAGAAGGAGAGACGAGAATGTTTGTCTTTAAGATGCG
CTGTTGCCGCTAACATGCGCTTTCGATTTTAGTGTAAACATGCGCCTCCATTGTTTTTG
GGTTTTGTTTTCGTCGTCGATAATCAAAGATTTTAAACACAATTCTCAAATTTTCACT
TGTTACAACTAGATTTGCATGATCTTTGTATTAACGAATAACGATTAAGTCCTAAA
>G24 Amino Acid Sequence (domain in AA coordinates: 25-93)
METEAAVTATVTAATMGIGTRKDLKPYKGIRMRKWGWAEIREPNKRSRIWLGSYATP
EAAARAYDTAVFYLRGPSARLNFPELLAGLTVSNGGGRGGLSAAIYIRRKAAEVGAQVDA
LGATVVVNTGGENRGDYEKIENCRKSGNSLERVDLNLKLPDENSDGDDDECVKRR*
>G2424 (1..999)
ATGAGGATGGAGATGGTGCATGCTGACGTGGCGTCTCTCTCCATAACACCTTGCTTCCCG
TCTTCTTTGTCTTCGTCTCATCATCACTATAACCAACAACAACATTGTATCATGTGCG
GAAGATCAACACCATTCGATGGATCAGACCACTTCATCGGACTACTTCTCTTTAAATATC
GACAATGCTCAACATCTCCGTAGCTACTACACAAGTCATAGAGAAGAAGACATGAACCCCT
AATCTAAGTGATTACAGTAATTGCAACAAGAAAGACACAACAGTCTATAGAAGCTGTGGA
CACTCGTCAAAAGCTTCGGTGTCTAGAGGACATTGGAGACCAGCTGAAGATACTAAGCTC
AAAGAACTAGTCGCCGTCTACGGTCCACAAAACCTGGAACCTCATAGCTGAGAAGCTCCAA
GGAAGATCCGGGAAAAGCTGTAGGCTTCGATGGTTTAAACCACTAGACCCAAGGATAAAT
AGAAGAGCCTTCACTGAGGAAGAAGAAGAGAGGCTAATGCAAGCTCATAGGCTTTATGGT
AACAAATGGGCGATGATAGTCGAGGCTTTCCCTGGTAGGACTGATAATTCTGTGAAGAAC
CATTGGCATGTTTATAATGGCTCGCAAGTTAGGGAACAATCTTCTTACCGTAGGAGG
AAGACGATGGTTTCTCTTAAGCCACTCATTAAACCCTAATCCTCACATTTTCAATGATTTT
GACCCTACCCGGTTAGCTTTGACCCACCTTGCTAGTAGTGACCATAAGCAGCTTATGTTA
CCAGTTCCTTGCTTCCAGGTTATGATCATGAAAATGAGAGTCCATTAATGGTGGATATG
TTCGAAACCCAAATGATGGTTGGCGATTACATTGCATGGACACAAGAGGCAACTACATTC
GATTTCTTAAACCAACCCGGAAGAGTGAGATATTTGAAAGAATCAATGAGGAGAAGAAA
CCACCATTTTTTCGATTTTCTTGGGTTGGGGACGGTGTGA
>G2424 Amino Acid Sequence (conserved domain in AA coordinates: 107-219)
MRMEMVHADVASLSITPCFPSSLSSSSHHYNNQQQHCIMSEDQHHSMDQTTSSDYFSLNI
DNAQHLSRYTSHREEDMNPNSLDYSNCKKDTTVYRSCGHSSKASVSRGHWPAEDTKL
KELVAVYGPQNWNLIAEKLQGRSGKSCRLRWFNQLDPRINRRAFTEEEERLMQAHRLYG
NKWAMIARLFPGRDTSVKNHWHVIMARKFREQSSSYRRRTMVSLLKPLINPNPHIFNDF
DPTRLALTHLASSDHKQLMLPVPCFPGYDHENESPLMVD MFETQMMVGDYIAWTQEATTF
DFLNQTKKSEIFERINEEKKPPFFDFLGLGTV*
>G2505 (1..1026)
ATGGGTTCTTCGTGCAACGGAGGAGTGCCACCTGGTTTCCGGTTTCATCCGACGGACGAA
GAGCTTCTCCATTACTACTTGAAGAAGAAAATCTCTTACCAAAAGTTTGAGATGGAAGTC
ATCAGAGAGGTTGACTTAAACAAGCTTGAGCCTTGGGATTTGCAAGAGAGATGCAAGATA
GGATCAACACCACAACGAATGGTACTTCTTCAGCCACAAGGACAGGAAATATCCGACG
GGGTCAAGGACCAACCGTGCTACTCATGCAGGGTTCTGGAAGGCGACGGGACGTGACAAG
TGCATAAGGAACTCTTACAAAAGATAGGAATGAGGAAGACACTTGTGTCTACAAAGGT
AGAGCTCCTCATGGCCAAAAGACTGATTGGATCATGCATGAGTACCGTCTTGAAGACGCT
GATGATCCTCAAGCCAACCTAGTGAAGATGGATGGGTGGTATGTAGAGTGTTTATGAAG
AAAAATTTGTTCAAGGTAGTAAATGAAGGTAGCTCAAGCATTAACTATTGGACCAACAC
AACCATGACGCATCTAACAACAACCATGCACCTCAAGCTCGTAGCTTTATGCACCGAGAC
AGTCCATACAGCTAGTACGTAAACACGAGCCATGACATTGAACTTAAACAAGCCTGAC
CTTGCTCTTCATCAATACCCACCAATCTTCCACAAGCCACCTTCACTTGGATTTGACTAC
TCTTCAGGACTTGCAAGGGACAGTGAGAGTGGCGCTAGTGAAGGGTTACAATACCAGCAA
GCGTGTGAGCCGGGTTTAGACGTTGGTACATGTGAGACAGTGGCTAGTCATAATCATCAA
CAAGGTCTAGGTGAATGGGCAATGATGGATAGACTTGTGACTTGTACATGGGAAATGAA
GATTCCTCTAGAGGGATTACGTATGAGGATGGTAACAACAATTCTGCTCTGTGGTTTCAG
CCAGTTCGCCGACGAACCAGCTAACATTGCGTAGTGAGATGGATTTCTGGGGTTATTCT
AAATAG
>G2505 Amino Acid Sequence (domain in AA coordinates: 10-159)
MGSSSNGGVPPGFRFHPTDEELLHYLKKKISYQKFEMEVIREVLDNLKLEPWDLQERCKI

GSTPQNEWYFFSHKDRKYPTGSRTNRATHAGFWKATGRDKCIRNSYKKIGMRKTLVIFYKG
RAPHGQKTDWIMHEYRLEDADDPQANPSEDGWVVCVFMKKNLFKVVNEGSSSINSLDQH
NHDASNNNHALQARSFMHRDSPYQLVRNHGAMTFELNKPDLALHQYPPIFHKPPSLGFDY
SSGLARDESAASEGLQYQACEPGLDVGTCETVASHNHQQGLGEWAMMDRLVTCHEMGNE
DSSRGITYEDGNNSSSSVVQPV PATNQLTLRSEMDFWGYSK*

>G2512 (64..798)

AACTTAGTGCCACTTAGACACAATAAGAAAACCGTTAACAAGAAGAAAAAAGATCG
AAAATGGAATATCAAACTAACTTCTTAAGTGGAGAGTTTTCCCGGAGAACTCTTCTTCA
AGCTCATGGAGCTCACAAGAATCATTCTTGTGGGAAGAGAGTTTCTTACATCAATCATT
GACCAATCCTTCCCTTTTATCTAGCCCTACTGATACTACTGTGATGACTTCTTGCATT
GAATCATCAATCATAAAAGAAGAAGGAAAAGAAGCCACCGTGGCGGCCGAGGAGGAGGAG
AAGTCATACAGAGGAGTGAGGAAACGGCCGTGGGGGAAATTCGCGGCCGAGATAAGAGAC
TCAACGAGGAAAGGGATAAGAGTGTGGCTTGGGACATTCGACACCGCGGAGGCGCGGCT
CTCGCTTATGATCAGGCGGCTTTCGCTTTGAAAGGCAGCCTCGCAGTACTCAATTTCCCG
GCGGATGTCGTTGAAGAATCTCTCCGGAAGATGGAGAATGTGAATCTCAATGATGGAGAG
TCTCCGGTGATAGCCTTGAAGAGAAAACACTCCATGAGAAACCGTCTAGAGGAAAGAAG
AAATCTTCTTCTTCTTCGACGTTGACATCTTCTCCTTCTTCTCCTCCTCTATTCTCT
TCTTCGTCTTCTTCTTCTTCTTGTGCTCAAGAAGTAGAAAACAGAGTGTGTTATGACGCAA
GAAAGTAATACAACACTTGTGGTCTTGTAGGATTTAGGTGCTGAATACTTAGAAGAGCTT
ATGAGATCATGTTCTTGATAATCTCTGCTTCTACAATTTTTATGTAATTTGA

>G2512 Amino Acid Sequence (conserved domain in AA coordinates: 79-139)

MEYQTNFLSGEFSPENSSSSSSWSSESFLWEESEFLHQSFQSFLLSSPTDNYCDDFFAFE
SSIIKEEGKEATVAEEEEESYRGVVRKRPWGKFAAEIRDSTRKGI RVWLGTDFDTAEAAAL
AYDQAAAFALKGSLAVLNFPADVVEESLRKMENVNLNDGESPVIALKRKHSRNRPRGKKK
SSSSSTLTSSPSSSSSYSSSSSSSSLSRSRKQSVVMTQESNTTLVVLEDLGAELYLEELM
RSCS*

>G2513 (69..698)

TTTCAACAGTAATTTAAGTTAACCGGAGTCTCTTTTGTGTTTCCGGCGAATTTTGGTAC
TTTGAGTTATGAATAATGATGATATTATTCTGGCGGAGATGAGGCCATAAGAAGCGTGCGG
GAAGGAGAGTGTTAAGGAGACACGTCACCCAGTTTACAGAGGCATAAGGCGGAGGAACG
GTGACAAATGGGTCTGCGAAGTCAGAGAACCGACGCACCAACGCCGCAATTTGGCTCGGGA
CTTATCCACAGCAGATATGGCAGCGCGTGACACGACGTGGCGGTTTTAGCTCTGCGTG
GGAGATCCGCATGTTTGAATTTCCGCCGACTCCGCTTGGCGGCTTCCGGTGCCGGAATCCA
ATGATCCGGATGTGATAAGAAGAGTTGCGGCGGAAGCTGCGGAGATGTTTAGGCCGTTGG
ATTTAGAAAGTGAATTACGGTTTTGCCTTGTGCGGAGATGATGTGGATTGGGTTTTG
GTTCCGGTTCCCGCTCTGGTTCCGGATCGGAGGAGAGGAATCTTCTTCGTATGGATTG
GAGACTACGAAGAAGTCTCAACGACGATGATGAGACTCGCGGAGGGGCCACTAATGTCCG
CGCCGCGATCGTATATGGAAGACATGACTCCTACTAATGTTTACACGGAAGAAGAGATGT
GTTATGAAGATATGTCATTGTGGAGTTACAGATATTAAGTGGGACTCACATATCTACTAT
ACATAATATTTAGCTTTTATGTAAAGGTATTTATGTGAGTTTTAAGATTGTAGATGTGT
CCAGGCGTTAGAAGTTTCTTGTGATGATGGAATCTTTGTACCTATAAAATTATAAAAT
T

>G2513 Amino Acid Sequence (domain in AA coordinates: TBD)

MNNDIILAEMRPKKRAGRRVFKETRHPVYRGIRRRNGDKWVCEVREPTHQRRRIWLGTYP
TADMAARAHDVAVLALRGRSACLNFADSAWRLPVPESNDPDVIRRVAAEAAEMFRPVDLE
SGITVLPAGDDVDLGFSGSGSGSGSEERNSSSYGFDYEEVSTTMMRLAEGPLMSPPR
SYMEDMTPTNVYTEEMCYEDMSLWSYRY*

>G2519 (83..691)

CAAAGTGAAAACATAAGATCATCTTCTTCGTTGATAGATCAATATAGGAACTCCAGAAGA
GAATCTTGATCAATTAAGTATCATGTCTCACATCGCTGTTGAAAGGAATCGAAGAAGGCA
AATGAACGAGCATCTTAAATCCCTTCGTTCTTTGACTCCTGTTTCTACATCAAAGGGG
AGATCAAGCTTCGATCATCGGAGGAGTGATAGAGTTTCAAAAGAGTTGCAGCAATTGGT
TCAAGTTCTTGAGTCCAAGAAACGTCGAAAGACCCTAAACCGACCATCTTCCCTTATGA
TCACCAGACAATCGAGCCATCCAGTTTAGGAGCCGCCACTACCCGAGTACCGTTTAGTCG
AATCGAAAATGTGATGACCACAAGTACTTTCAAGGAAGTAGGAGCATGCTGTAACCTCCC
TCATGCTAACGTAGAAGCAAAGATTTCAAGTTCTAATGTTGTATTGAGAGTTGTCTCTAG
CGGAATCGTGGGCGAGCTCGTAAAGATCATCTCTGCTTTAGAGAAGCTATCTTTCAAGT

TCTTCACCTCAATATTAGTAGCATGGAGGAGACTGTCTTATACTTTTTTCGTTGTTAAGAT
AGGATTGGAGTGTCACTTAAGCTTGGAGGAGCTAACTCTTGAAGTTCAGAAAAGCTTTGT
GTCTGATGAAGTGATCGTCTCTACCAATTA AAAACAAAATCTACATGTACTAGAGCGTG
TATCGTTTTTTGGGATTAATAATCATATAATCGTTACATGAGCCTTGATACTTTGCTAGA
AATAAGCTCCTCTAAACAAAACCTTCTTTTTAAAAAACACACTTATGTTTTACTTAGTT
TGTTGTTGTATCCGAAGTTGATCAACGTTGTAATTTCCCAATAAATCATGACATTTTA
TATGCTCT

>G2519 Amino Acid Sequence (domain in AA coordinates:1-65)

MSHIAVERNRRRQMEHLKSLRSLTPCFYIKRGDQASIIGGVIEFIKELQQLVQVLESKK
RRKTLNRPSFPYDHQTIIEPSSLGAATTRVPFSRIENVMTTSTFKEVGACCNSPHANVEAK
ISGSNVVLRVVSRRIVGQLVKIISVLEKLSFQVLHLNISSMEETVLYFFVVKIGLECHLS
LEELTLEVQKSFVSDEVIVSTN*

>G2520 (133..1197)

AAGGAGTTTTGCATACTACCAAGCCACAATCATTTCTCTCTCTATCTCTCTGGTTT
TGAATCGGCGACGACTGAGTCAACTCGGTGTTGTTACTGGTTTCGTCGTATGTGTTGTAA
CTGATTAAGTTGATGGATCCGAGTGGGATGATGAACGAAGGAGGACCGTTTAATCTAGCG
GAGATCTGGCAGTTTCCGTTGAACGGAGTTTCAACCGCCGAGATTCTTCTAGAAGAAGC
TTCGTTGGACCGAATCAGTTCGGTGATGCTGATCTAACCACAGCTGCTAACGGTGATCCA
GCGCGTATGAGTCACGCGTTGTCTCAGGCGGTTATTGAAGGTATCTCCGGCGCTTGGAAA
CGGAGGGAAGATGAGTCTAAGTCGGCGAAGATCGTCTCCACCATTGGCGCTAGTGAAGGT
GAGAACAAAAGACAGAAGATAGATGAAGTGTGTGATGGGAAAGCAGAAGCAGAATCGCTA
GGAACAGAGACGGAACAAAAGAAGCAACAGATGGAACCAACGAAAGATTATATTCATGTT
CGAGCTAGAAGAGGTCAAGCTACTGATAGTCACAGTTTAGCTGAAAGAGCGAGAAGAGAG
AAAATAAGTGAGCGGATGAAAATCTTGCAAGATCTTGTTCCGGGATGTAACAAGGTTATT
GGAAAAGCACTTGTTCTAGATGAGATAATTAATATATACAATCATTGCAACGTCAAGTT
GAGTTCTTATCGATGAAGCTTGAAGCAGTCAACTCAAGAATGAACCCTGGTATCGAGGTT
TTTCCACCCAAAGAGGTGATGATTCTCATGATCATCAACTCAATCTTCTCCATTTTTTTC
ACAAAACAATACATGTTTCTATCGAGGTATTCTCGGGGTAGGAGTCTCGATGTTTATGCG
GTTTCGGTCATTTAAGCATTGCAATAAACGGAGTGACCTCTGTTTTTGTCTCTGCTCCCCA
AAAACAGAACTTAAGACAATATATTTTACAAAACATGACATGTTTCTGTGATATTCT
CGAGTAGGAGTCGCTATTAGTTCATCTAAGCATTGCAATGAACCGTTTGGTCAGCAAGCG
TTTGAGAATCCGGAGATACAGTTCGGGTCGCAGTCTACGAGGGAATACAGTAGAGGAGCA
TCACCAGAGTGTTGTCACATGTCAGATAGGATCAGGTGGTTTCGAAAGAACGTCTTGA

>G2520 Amino Acid Sequence (domain in AA coordinates: 135-206)

MDPSGMMNEGGPFNLAEIWQFPLNGVSTAGDSSRRSFVGPNOFGDADLTAAANGDPARMS
HALSQAVIEGISGAWKRREDESKSAKIVSTIGASEGENKRQKIDEVCDGKAEBSLGTET
EQKKQQMEPTKDYIHVRARRQATDSHSLAERARREKISERMKILQDLVPGCNKVIKAL
VLDEIINYIQLQRQVEFLSMKLEAVNSRMNPGIEVFPPKEVMILMIINSIFSIFFTKQY
MFLSRYSRGRSLDVYAVRSFKHCNKRSDLCFCSCSPKTELKTTIFSQNMTCFCRYSRVGV
AIISSSKHCNEFPFQAFENPEIQFGSQSTREYSRGASPEWLHMQIGSGGFERTS*

>G2533 (1..1080)

ATGATAAGCAAGGATCCAATATCGAGTTTACCTCCAGGGTTTCGATTTTCATCCAACAGAT
GAAGAACTCATTCTCCATTACCTAAGGAAGAAAGTTTCTCTTCCCCAGTCCCGCTTTCG
ATTATCGCCGATGTGATATCTACAAATCCGATCCATGGGATTTACCAGCTAAGGCTCCA
TTTGGGGAGAAAGAGTGGTATTTTTCAGTCCGAGGGATAGGAAATATCCAAACGGAGCA
AGACCAAACAGAGCAGCTGCGTCTGGATATTGGAAGCAACCGGAACAGATAAATTGATT
GCGGTACCAAATGGTGAAGGGTTTCATGAAAACATTGGTATAAAAAAGCTCTTGTGTTT
TATAGAGGAAGCCTCCAAAAGGTGTTAAACCAATTGGATCATGCATGAATATCGTCTT
GCCGATTCAATATCTCCAAAAGAAATTAACCTCTTCTAGGAGCGGTGGTAGCGAAGTTAAT
AATAATTTTGGAGATAGGAATCTTAAAGAAATATTGATGAGACTGGATGATTGGGTCTT
TGCCGGATTTACAAGAAATCACACGCTTCATTGTATCACCTGATGTTGCTTTGGTCACA
AGCAATCAAGAGCATGAGGAAAATGACAACGAACCATTCGTAGACCGCGGAACCTTTTTCG
CCAAATTTGCAAAATGATCAACCCCTTAAACGCCAGAAGTCTTCTTGTTCGTTCTCAAAC
TTACTAGACGCTACAGATTTGACGTTTCTCGCAAATTTTCTAAACGAAACCCCGGAAAAT
CGTTCTGAATCAGATTTTCTTTTCATGATTGGCAATTTCTCTAATCCTGACATTTACGGA
AACCATTACTTGGATCAGAAGTTACCGCAGTTGAGCTCTCCCACTTCAGAGACAAGCGGC
ATCGGAAGCAAAAGAGAGAGAGTGGATTTTTCGGAAGAAACGATAAACGCTTCGAAGAAG

ATGATGAACACATATAGTTACAATAATAGTATAGATCAAATGGATCATAGTATGATGCAA
 CAACCTAGTTTCCTGAACCAGGAACCTCATGATGAGTTCTCACCTTCAATATCAAGGCTAG
 >G2533 Amino Acid Sequence (conserved domain in AA coordinates:11-186)
 MISKDPISSLPPGFRFHPTEDEELILHYLRKKVSSSPVPLSIIADVDIYKSDPWDLPAKAP
 FGEKEWYFFSPDRKYPNGARPNRAASGYWKATGTDKLIAPNPEGFHENIGIKKALVF
 YRGKPPKGVKTNWIMHEYRLADSLSPKRINSSRSGGSEVNNFNGDRNSKEYSMRLDDWVL
 CRIYKKSHASLSSPDVALVTSNQEHEENDNEPFVDRGTFPLNLQNDQPLKRQKSSCSFSN
 LLDATDLTFLANFLNETPENRSESDFSFMIGNFSNPDYGNHYLDQKLPQLSSPTSETSG
 IGSKRERVDFAETINASKKMMNTYSYNNSIDQMDHSMMQPSFLNQELMMSSHLQYQG*
 >G2534 (1..975)

ATGGATAATATAATGCAATCGTCAATGCCACCGGGATTCCGATTTCATCCGACAGAGGAA
 GAGCTTGTGGGTTTACCTAGATAGGAAGATCAATTCAATGAAGAGTGCTTTAGATGTC
 ATTGTAGAGATTGATCTCTACAAAATGGAGCCATGGGATATACAAGCGAGGTGTAAACTA
 GGGTATGAAGAGCAAAACGAGTGGTACTTCTTTAGTCATAAGGACAGGAAGTACCCTACC
 GGGACTAGGACCAACCGAGCCACTGCGGCTGGGTTCTGGAAAGCCACGGGTAGAGACAAG
 GCGGTACTATCAAAAAACAGTGTGCATCGGAATGCGGAAGACACTTGTCTACTACAAGGGT
 CGAGCTCCTAATGGAAGAAAGTCCGATTGGATCATGCACGAATACCGTCTCCAAACTCC
 GAGCTTGCCCCGGTTCCAGGAGGAAGGCTGGGTGGTGTGTCGAGCATTTAGGAAGCCAATT
 CCAAACAGAGGCCATTAGGGTACGAGCCATGGCAGAACCAGCTCTACCACGTCGAAAGT
 AGTAACAACACTACTCATCTTCAGTGACAATGAACACGAGTCATCATATCGGTGCATCTTCA
 TCAAGTCATAACCTTAATCAAATGCTCATGAGCAATAACCACTACAATCCTAATAATACA
 TCCTCATCGATGCATCAATATGGCAACATTGAGCTCCCGCAGTTGGACAGCCCCGAGCTTG
 TCGCCTAGTTTAGGGACGAATAAAGATCAGAACGAGAGTTTCGAGCAAGAAGAAGAGAAG
 AGCTTTAACTGTGTGGATTGGAGAACACTAGATACCTTGCTTGAGACACAAGTCATACAT
 CCGCATAACCCTAATATTCTTATGTTGAAACGCAGTCGTATAATCCGGCGCCAGCTTC
 CTTTCCATGCATCAAAGCTATAATGAGGTCGAAGCTAATATTCATCATTCTCTTGATGC
 TTCCTGACTCGTAA

>G2534 Amino Acid Sequence (conserved domain in AA coordinates:10-157)
 MDNIMQSSMPGFRFHPTEEBELVGYLDKINSMSALDVIVEIDLYKMEPWDIQRCKL
 GYEEQNEWYFFSHKDRKYPTGTRTNRATAAGFWKATGRDKAVLSKNSVIGMRKTLVYYKG
 RAPNGRKS DWIMHEYRLQNSLAPVQEEGWVVCRAFRKPIPNQRPLGYEPWQNQLYHVES
 SNNYSSSVTMTNTHHIGASSSSHNLMMSNNHYNPNNTSSSMHQYGNIELPQLDSPSL
 SPSLGTNKDQNESFEQEEKSFNCVDWRTLDTLLETQVIHPHNPNI LMFETQSYPAPSF
 PSMHQSYNEVEANIHHS LGCFPDS*

>G2573 (34..957)
 CCAGATTTAATTTGAGACTCTCAAAGAAACACCATGGAAGAAGAGCAACCTCCGGCCAAG
 AAACGAAACATGGGGAGACTCTAGAAAAGGTTGCATGAAAGGTAAAGGCGGTCCAGAGAAC
 GCCACGTGTACTTTCCGTGGAGTTAGGCAACGGACTTGGGGTAAATGGGTGGCTGAGATC
 CGTGAGCCTAACCGTGGGACTCGTCTCTGGCTCGGCACGTTTAATACCTCGGTCCGAGGCC
 GCCATGGCTTACGATGAAGCCGCTAAGAACTCTATGGACACGAGGCTAAACTCAACTTG
 GTGCACCCACAACAACAACAAGTAGTAGTGAACAGAACTTGTCTTTTCTGGCCAC
 GGGTCGGGTTCTTGGGCTTATAATAAGAAGCTCGATATGGTTTCATGGGTTGGACCTTGGT
 CTCGGCCAGGCAAGTTGTTACAGAGGTTCTTGCTCAGAGAGATCGAGTTTTCTACAAGAA
 GATGATGATCATAGTCATAATCGATGTTTCGTCTTCAAGTGGTTTCAATCTTTGTTGGTTA
 TTACCTAAACAAAGTGATTACAAAGATCAAGAGACCGTTAATGCTACGACTAGTTATGGC
 GGTGAAGGCGGTGGTGGCTTACGTTAACGTTTTTCGACCAATTTGAAACCAAGAATTG
 ATGAGTCAGAATTATGATTATACAATGGAGCTTGGTCTAGGTTTCTTGTGGGGCAAGAA
 AAGAAGACGGAACATGACGTGTCTCGTGTGGATCGTCGGACAACAAGGAGAGTATG
 TTGGTTCTAGTTGCGGCGGAGAGAGGATGCATAGCCGGAGTTGGAAGAGCGAACAGGA
 TATTTGGAATGGATGATCTTTTGGAGATTGATGATTAGGTTTGTGATTGGCAAAAAT
 GGAGATTTCAAGAATTGGTGTGTGAAGAGTTTCAACATCCATGGAATTGGTTCTGAGAG
 TTTTATTTATTACTATTATTTATCATACATATTTCTTATATTTGACTTAGG

>G2573 Amino Acid Sequence (domain in AA coordinates: TBD)
 MEEEQPPAKKRNMGSRKGMKGKGGPENATCTFRGVQRQTWGWVABIREPNRGRTRLWL
 GTFNTSVEAAMAYDEAAKLYGHEAKLNLVHPQQQQQVNVNRNLSFSGHSGSGSWAYNKKL
 DMVHGLDLGLGQASCSRGSCSERSSFLQEDDDHSHNRCSSSSGSNLCWLLPKQSDSQDQE
 TVNATTSYGGEGGGSTLTFSTNLKPNLMSQNYGLYNGAWSRFLVGQEKTEHDVSSSC

GSSDNKESMLVPSGGERMHRPELEERTGYLEMDDLLEIDDLGLLIGKNGDFKNWCCEEF
QHPWNWF*

>G2589 (23..1354)

AAAGAAAAGAAAAATAAAGATAATGAGGACGAAGACTAAGTTAGTACTCATACCTGATAG
ACACTTTCGGAGAGCCACATTCAGGAAGAGGAATGCAGGGATAAGGAAGAACTCCACGA
GCTGACAACTCTCTGTGACATCAAAGCATGTGCGGTAATCTACAGTCCGTTTCGAGAATCC
AACGGTGTGGCCGTCACCCGAAGGTGTTCAAGAGGTGATTTTCGGAGTTTCATGGAGAAGCC
GGCGACAGAACGGTCCAAGACGATGATGAGTCATGAGACTTTCTTGCGGGACCAAATCAC
CAAAGAACAAACAACTAGAGAGTCTACGTCGTGAAAACCGAGAACTCAGCTTAAGCA
TTTTATGTTTGATTGCGTTGGAGGCAAGATGAGTGAGCAACAGTATGGTGCAAGGGACCT
TCAAGATTTAAGTCTTTTTACTGATCAATATCTTAATCAGCTTAATGCCAGGAAGAAGTT
CCTTACAGAATATGGTGAGTCTTCTTCTTCTGTTCTCTCTGTTTGATGTTGCGGGTGC
CAATCTCCTGTTGTTGTCAGATCAAGCTGCGGTAAGTGTCTCTCTTGTGTTGCTGTTGC
GGGTGCCAATCTTCTGTTGTTGCTGATCAAGCTGCGGTAAGTGTCTCTCTGTTTGC
TGTTGCGGGTGCCAATCTTCTGTTGTTGCTGATCAAGCTGCGGTAAGTGTCTCTCTGTTTGC
ATTTTCATAACATGAATGTGAACCAAGATCAGTATGAGCCGGTTCAGCCCTATGTCCCTAC
TGTTTTTAGTGATCATATTCAATATCAGAATATGAACCTCAATCAAAACCAACAAGAGCC
GGTTCATTACCAGGCTCTTGCTGTTGCGGGTGCCGGTCTTCTATGACTCAGAATCAGTA
TGAGCCCGTTCACTACCAGAGTCTTGCTGTCGCGGGTGCGGGTCTTCTATGAGTCAGTT
GCAGTATGAGCCGGTTCAGCCTTATATCCCTACTGTTTTTAGTGATAATGTTCAATATCA
GCATATGAATTTGTATCAAAATCAACAAGAGCCGGTTCCTACCAAGCTCTTGGTGTTGC
AGGTGCCGGTCTTCTATGAATCAGAATCAGTATGAGCCGGTTCAGCCCTATGTCCCTAC
TGGTTTTAGTGATCATTTTTAGTTTGAGAAATGAATTTGAATCAAAATCAACAGGAGCC
GGTTCAATACCAAGCTCTCTGTTGATTTTAATCATCAGATTCAACAAGGAACTATGATAT
GAATTTGAACCAAGATATGAGTTTGATGTCAAATCAGTATCCGTTTCAAAATGATCCATT
CATGAATATGTTGACAGAATATCCTTATGAATAAGCGGGTTATGTTGGAGAGCATGCAC

>G2589 Amino Acid Sequence (domain in AA coordinates: TBD)

MRTKTKLVLPDRHFRRATFRKRNAGIRKKLHELTTLCDIKACAVIYSPFENPTVWPSTE
GVQEVISEFMEKPAERSKTMMSHETFLRDQITKEQNKLESLRRENRETQLKHFMFDCVG
GKMSEQQYGARDLQDLSLFTDQYLNQLNARKKFLTEYGESSSSVPLFDVAGANPPVVAD
QAAVTVPPLFAVAGANLPVVADQAAVTVPPLFAVAGANLPVVADQAAVNVPTGFHNMNVN
QNQYEPVQPYVPTGFSDDHIQYQNMNFNQNQEPVHYQALAVAGAGLPMTQNQYEPVHYQS
LAVAGGGLPMSQLQYEPVQPYIPTVFSNVQYQHMNLQYQEPVHYQALGVAGAGLPMN
QNQYEPVQPYVPTGFSDDHFQFENMNLNQNQEPVQYQAPVDFNHQIQQGNYDMNLNQNMS
LDPNQYPFQNDPFMNLTEYPYE*

>G2687 (45..1139)

CTCTGTCTCTCGTATCTTTCTACTACTCTGTTTCTTGAATTCTAATGAACAACATCGACG
ACGCAAAGACGGAGACTTCAGTGTCTTCAGGTTCAAGCGACTCTTTCTTGCTCTCAAGA
AACGCATGAGACTTGATGACGAACAGAAACGCCCTAGTGGTTTCGTCTTCACCAAAGA
CGGTTGTGGCTTCTGGCAATGTCAAGTACAAAGGAGTCGTTACGCAACAGAACGGTCATT
GGGGTGCCAGATTTACGCAGACCACAAAAGGATTTGGCTTGGAACCTTCAAATCCGCTG
ATGAAGCCGCCACGGCTTACGATAGTGCATCTATCAAACCTCCGAAGCTTTGACGCTAACT
CGCACCGGAACCTCCCTTGGTCTACAATCACTCTCAACGAACAGACTTTCAAAATTGCT
ACACAACAGAGACTGTGTTGAACATGATCAGAGACGGTTCGTACCAACACAAATTCAGAG
ATTTTCTCAGAATCAGATCTCAGATTGTTGCGAGTATCAACATCGGGGGACCAAAACAAG
CCCGAGGAGAAGTGAATCAAGAATCAGACAAGTGTTTTTCTTGACACAGCTTTTTCAGA
AGGAATTGACACCGAGCGATGTAGGGAACTAAATAGGCTTGTGATACCTAAAAAGTATG
CAGTGAAGTATATGECTTTTCATAAGCGCTGATCAAAGCGAGAAAGAAGAGGGTGAAATAG
TAGGATCTGTGGAAGATGTGGAGGTTGTGTTTTACGACAGAGCAATGAGACAATGGAAGT
TTAGGTATTGTTACTTGGAAAAGTAGCCAGAGCTTGTCTTACCAGAGGATGGAATAGTT
TCGTGAAGGAGAAGAATCTCAAGGAGAAGGATGTTATTGCCTTCTACACTTGCGATGTCC
CGAACATGTGAAGACATTAGAAGGTCAAAGAAAGAACTTCTTGATGATCGATGTTTCATT
GCTTTTCAGACAACGGTTCCGTGGTAGCTGAGGAAGTAAGTATGACGGTTCATGACAGTT
CAGTGCAAGTAAAGAAAACAGAAAACCTTGGTTAGCTCCATGTTAGAAGATAAAGAAACCA
AATCAGAGGAGAACAAAGGAGGGTTTATGCTGTTTGGTGTAAGGATCGAATGTCCTTAGG
GAATTTTTCTTTAAAAGTTTCTTACTTCACTAGAACTTGTTTTACTTGTACCT

>G2687 Amino Acid Sequence (domain in AA coordinates: TBD)

MNNIDDAKTETSVSSGSSDSFLPLKKRMLDDEPENALVVSSSPKTVVASGNVYKGVVQ
QQNGHWGAQIYADHKRIWLGTFKSADEAATAYDSASIKLRSFDANSHRNPWFSTITLNEP
DFQNCYTTETVLNMIRDGSYQHKFRDFLRIRSQIVASINIGGPKQARGEVNQESDKCFSC
TQLFQKELTPSDVGKLNRLVIPKKYAVKYMPPFISADQSEKEEGEIVGSVEDVEVVFYDRA
MRQWKFRYCYWKSSQS FVFTRGWNSFVKEKNLKEKDVI AFYTC DVPNNVKTLEGQRKNFL
MIDVHCFSDNGSVVAEEVSM TVHDS SVQVKKTENLVSSMLEDKETKSEENKGGFMLFGVR
IECP*

>G27 (83..622)

CAAAATACCAAAAACAAAACATTTTTTTTAATCTTCCCACCAATTTTTTCTCTTTCTCT
CGTTACATTAAATTATCTTTAGATGCAAGACTCTTCTCTCACGAATCGCAACGTAACCT
CCGGTACACCGGTGCCGGAGAAAACCGGAAAGAGTTCTAAGACTAAAAATGAGCAAAAAGG
TGTTTCTAAACAACCAAAATTTTCGTGGGGTCAGAATGAGACAATGGGGAAAATGGGTGTC
TGAAAATTAGAGAACCAAGAAAGAAATCAAGAATATGGCTCGGTACTTTCTCTACGCCGGA
GATGGCGGCGCGTGCACACGACGTGGCGGCTTTAGCCATCAAAGGTGGCTCTGCCACCT
TAATTTCCCGGAGCTAGCTTACCATTGCGGAGACCGGCTAGCGCGGACCCTAAAGACAT
TCAAGAAGCCCGCCGCCGAGCAGCTGCCGTTGACTGGAAAGCACCGGAGTCTCCGTCTAG
CACCGTGACGTCTCTCAGTCGCCGACGACGCTTCTCCGATCTTCTGATCTTTTGCT
TGACGTGAATGATCACAACAAAACGATGGATTCTGGGACTCGTTTCCGTACGAAGATCC
TTTCTTCTTGAAAATTACTAGAAGGCAAATCTTGCCGGCGAAGCGATTTTCCGGTGGT
TTCCCGGTAAATAAGAAGACGATGTGCTTTGTACCTTTTGTCTACGATGGGAAATTT
CTTTTCTTTTACGTGTAGTAAAGTTTCCGAATGTGTGATGTGTAAGTAAGTACAGGT
TATTTAATTTCTTTTCTTTTGTACAAATACGTACGTATTACCAAAAAGTTTTCATTTATT
GTGCTTTTATCTTCCAAATTCATTAAAAAAAAAAAAAAAAAA

>G27 Amino Acid Sequence (domain in AA coordinates: 37-104)

MQDSSSHESQRNLRSPVPEKTGKSSKTKNEQKGVSKQPNFRGVRMRQWGWVSEIREPRK
KSRIWLGTFTSTPEMAARAHDAALAIKGGSAHLNFPDELAYHLPRPASADPKDIEAAAAA
AAVDWKAPESPSSTVTSSPVADDAFSDLPLDLLLDVNDHNKNDGFWDSPFYEDPFFLENY*

>G2720 (1..894)

ATGGAAGCGAAGAAGAGATAAAGAAAGGTCCATGGAAAGCCGAAGAAGACGAAGTA
CTCATCAACCATGTCAAGAGATACGGTCTCGTGATTGGAGCTCCATTCGATCCAAAGGT
CTTCTTCAACGCACCGGCAAATCCTGTCTCTCGTTGGGTCAATAAACTCCGTCCCAAT
CTCAAAAATGGATGCAAGTTCTCGGCTGACGAAGAGAGGACTGTGATTGAGTTACAATCT
GAGTTTGGAACAAATGGGCGAGAATCGCTACGTATCTACCGGGAAGAACTGATAACGAT
GTGAAGAATTTCTGGAGTAGCAGACAAAAGAGACTCGCTAGGATTCTTCATAACTCCTCT
GATGCATCGAGTTTCGAGTTTCAATCCCAAATCTTCTTCTTCTCATCGACTCAAGGGCAAA
AACGTCAAACCAATCCGTCAATCCTCTCAGGGTTTTGGTTTTGGTTGAGGAAGAGGTTACA
GTTTCTTCTTCAATGTTCCAGATGGTTCCTTATTCATCTGATCAAGTTGGTGATGAAGTC
TTGAGGTTGCCGGATTGCGGTGTTAAGTTAGAGCATCAGCCTTTTCGCTTTTGGCACTGAT
CTTGCTCTAGCAGAGTACTCTGACTCACAGAATGATGCAATCAGCAAGCAATCAGCCCT
TTCTCTCCAGAAAGCAGAGAGCTTTTGGCTAGACTTGACGACCCTTTTACTATGATATA
CTTGGACCAGCTGATTCTTCTGAGCCATTGTTGCTCTCCCTCAGCCGTTCTTCGAGCCT
TCGCCTGTGCCGAGAAGATGCAGACATGTTTCAAAGGATGAAGAAGCTGATGTTTCTTA
GACGATTTCCAGCTGACATGTTTGATCAGGTTGATCCAATCCCAAGTCCTTAG

>G2720 Amino Acid Sequence (domain in AA coordinates: 10-114)

MEAKKEEIKKGPWKAEEDEVLINHVKRYGPRDWSSIRSKGLLQRTGKSCRLRWVNLKLRPN
LKNGCKFSADERTVIELQSEFGNKWARIATYLPGRTDNDVKNFWSRQKRLARILHNS
DASSSSFNPKSSSSSHRLKGNVKPIRQSSQGFGLVEEVEVTSSSSCSQMVPSYSDQVGVDEV
LRLPDLGVKLEHQPFAGTDLVLAEYSDSQNDANQQAISPFSPEARELLARLDDPFYDI
LGPADSSEPLFALPQFFPEPSPVPRRCRHVSKDEEADVFLDDFPADMFQVDPPIPS*

>G2787 (142..1584)

TCTCAGAGCAAAAAACAAAAAAGAAAAAACCCTAAATCTAAATCTCACCTTCCA
CCTCTGTCTTTTTTTTTTTTGTCTTTTTTTTTTTTTTACTGTATCTTCTCTTCTTTG
CTCTGCAAAAATCTCACATCCATGGATCCATCTCTTGGTGATCCTCATCATCTCCTCAG
TTCACCCCTTTTCTCATTTTCCACCTCCAATCATCATCTTTAGGACCAATCCGTAC
AATAACCATGTCCTCTTCCAACCGCAGCCGCAACGCAACGCAATCCCGCAACCGCAG
ATGTTTCAGTTATCTCCACATGTTTCAATGCCCCACCCTTCTACTCCGAAATGATTGTC
GCTGCGATTGCGGCGTTAAACGAACCGGATGGTTGAGCAAGATGGCAATTCGAGATAC

ATCGAGAGATGTTACACCGGTTTAACTTCTGCTCATGCTGCTTTGTTGACTCACCATCTC
 AAGACTTTGAAGACCAGTGGTGTCTTTCTATGGTTAAGAAATCTTACAAAATTGCTGGT
 TCTTCTACTCCTCCTGCTAGTGTAGCTGTTGCTGCTGCTGCCGCCGCTCAAGGTCTCGAT
 GTTCCCAGATCTGAGATTCTCCATTCAAGTAACAACGATCCCATGGCTTCTGGCTCTGCT
 TCTCAGCCTCTGAAACGAGGTCTGGTGTCTCTCTAAGCCTAAACCTGAATCTCAACCA
 CAACCACTACAGCAACTTCCACCGACCAATCAAGTCCAGGCTAACGGACAGCCAATCTGG
 GAACAGCAGCAAGTTCAATCACCTGTTCGGTTCCGACTCCGGTTACAGAGTCGGCGAAG
 AGAGGACCTGGTCTGCTCAAGGAAGAACGGTTCTGCTGCTCTGCTACTGCACCAATCGTT
 CAAGCTTCGGTTATGGCTGGAATTATGAAACGTAGAGGTAGACCACCGGGTCTGCGAGCT
 GCTGGGAGACAGAGGAAGCCCAAATCCGTTTCTTCTACTGCCTCTGTGTATCCTTATGTT
 GCTAATGGTGTAGACGCAGAGGAAGGCCTAGGAGAGTTGTTGACCCTAGCAGTATTGTT
 AGTGTGTGCTCCAGTAGGTGGTGAATGTGGCAGCGGTTGCGCCAGGGATGAAGCGTGGA
 CGTGGACGACCACCTAAGATTGGTGGTGTATCAGTAGGCTTATTATGAAGCCTAAGAGA
 GGACGAGGACGCTCTGTAGGTAGACCCAGAAAGATTGGAACATCAGTCACGACTGGGACA
 CAAGATTCTGGAGAACTCAAGAAGAAGTTTGATATTTTCAAGAGAAAGTGAAGAAATT
 GTGAAGGTGTTGAAGGATGGAGTTACAAGTGAGAATCAAGCAGTGGTGCAAGCCATAAAA
 GATCTGGAAGCACTAACAGTGACGGAGACCGTTGAGCCACAAGTTATGGAAGAAGTGCAG
 CCAGAGGAGACTGCAGCACCACAGACTGAAGCTCAACAACTGAAGCTGCTGAGACACAA
 GGAGGACAAGAAGAAGGACAAGAAAGAGAAGGAGAAACACAGACCCAGACAGAAGCAGAG
 GCAATGCAAGAAGCTCTGTTCTGAAGAATAATAATGATCTAGAAAACAACCTAGACATAA
 TAGCCTTGGTGTGTTGGCGTTAGGAGTGTGTTTTTTTTTAGTTGTTTAGGTGTTGGAATCGC
 ATCTTAAATTATATAAAAATCTATAAGGAATTTTAAATTTTCTAGGTTTTGTTGTCTGCA
 GAAGAAGAAATAGTAGACTCGTTAATGGTGTGTTGTGCGGTGTGCTTTAACCAAACCAT
 AAGACGTGGCTGTAAATTAGCGATGTTTCTAGTCTTCCATCTTTAATAATCTCTTATTGC
 GTCTGTGCTTTGTTTTT

>G2787 Amino Acid Sequence (domain in AA coordinates: 172-192, 226-247, 256-276, 290-311, 245-366)

MDPSLGDPHHPQFTFPFPHFPTSNNHPLGPNPYNNHVVFQPPQQTQTQIPQPQMFQLSPH
 VSMPPHPYSEMICAIAALNEPDGSSKMAISRYIERCYTGLTSAHAALLTHHLKTLKTSG
 VLSMVKKSYKIAGSSTPPASVAVAAAAAQGLDVPRSEILHSSNNDPMASGSASQPLKRG
 RGRPPKPKPESQPQPLQQLPPTNQVQANGQPIWEQQQVQSPVPVPTPVTESAKRGPGRPR
 KNGSAAPATAPIVQASVMAGIMKRRGRPPGRRAAGRQRKPKSVSSTASVYPYVANGARRR
 GRPRRVDPSSIVSVAPVGGENVAAVAPGMRGRGRPPKIGGVISRLIMKPKRGRGRPVG
 RPRKIGTSVTTGTQDSGELKKKFDIFQEKVKEIVKVLKDGVTSENQAVVQAIKDLEALTV
 TETVEPQVMEEVQPEETAAPQTEAQOTEAAETQGGQEEGQEREGETQTQTEAEAMQEALF
 *

>G2789 (82..879)

CTTTAGGGACACCAAATCTATTCAACCTAAAAGCCTTCTTTTCCCTATATTGACCAACT
 TTTTAGCGAATCAGAAGAGGAATGGATGAGGTATCTCGTTCTCATACACCGCAATTTCTA
 TCAAGTGATCATCAGCACTATCACCATCAAAACGCTGGACGACAAAAACGGGCAGAGAA
 GAAGAAGGAGTTGAACCAACAATATAGGGGAAGACCTAGCCACCTTTCCTTCCGGAGAA
 GAGAATATCAAGAAGAGAAGGCCACGTGGCAGACCTGCTGGTTCCAAGAACAACCCAAA
 GCACCAATCATAGTCACTCGCGACTCCGCGAACGCCTTCAGATGTCACGTATGGAGATA
 ACCAACGCCTGCGATGTAATGGAAAGCCTAGCCGTCTTCGCTAGACGCCGTAGCGTGGC
 GTTTGCGTCTTGACCGGAAACGGGGCCGTTACAAACGTCACCGTTAGACAACCTGGCGGA
 GGCGTCGTGAGTTTACACGGACGGTTTGAGATTCTTTCTCTCTCGGGTTCGTTTCTCCT
 CCACCGGCACCACAGCTGCGTCTGGTTTAAAGGTTTACTTAGCCGGTGGTCAAGGTCAA
 GTGATCGGAGGCAGTGGTGGGACCGCTTACGGCATCAAGTCCGGTGGTGGTTATGGCA
 GCTTCATTTGAAACGCATCTTACGAGAGGCTGCCACTAGAGGAGGAGGAGAACTGAA
 AGAGAAATAGATGGAAACGCGGCTAGGGCGATTGGAACGCAACGCAGAAACAGTTAATG
 CAAGATGCGACATCGTTTATTGGGTGCGCGTCGAATTTAATTAACCTCTGTTTCGTTGCCA
 GGTGAAGCTTATTGGGGAACGCAACGACCGTCTTTCTAAGATAATATCATTGATAATATA
 AGTTTCGTCTTCTATTCTTTTCACTTTTACCTTTTCACTTTCTTAGGTTTTGTTTT
 AACGTTTGATTAATACCTGAAGGTTTTTGAAAATTTTCGATCGGATAAAGGATTTATG
 TTGCGAGCCGAAACGCGGCC

>G2789 Amino Acid Sequence (domain in AA coordinates: 53-73, 121-165)

MDEVSRSHTPQFLSSDHQHYHHQNAQRQKRGREEGVPEPNIGEDLATFPSGEENIKRR

PRGRPAGSKNPKAPIIVTRDSANAFRCHVMEITNACDVMEISLAVFARRRQRGVCVLITGN
GAVTNVTVRQPGGGVVSLHGRFEILSLSGSFLPPPAPPAASGLKVYLAGGQGVIGGSV
GPLTASSPVVMAASFGNASYERLPLEEEEEETEREIDGNAARAIGTQTQKQLMQDATSFI
GSPSNLINSVSLPGEAYWGTQRPSF*

>G31 (13..615)

CTTTTATAAGCAATGGCTCCAAGACAGGCGAACGGTAGAAGCATTGCCGTGAGTGAAGGC
GGCGGAGGGAAGACGATGACGATGACGACGATGCGGAAGGAAGTGCACCTTAGAGGTGTG
AGGAAGCGTCCATGGGGTAGATACGCGGCGGAGATCCGTGACCCGGGAAAGAAAACCCGG
GTTTGGCTCGGGACATTTCGACACGGCGGAGGAAGCTGCAAGAGCTTACGACACCGCCGCT
AGAGAGTTTCGTGGCTCCAAAGCAAAGACTAATTTCCCTCTTCCCGGAGAGTCTACTACG
GTTAACGACGGTGGCGAGAACGATTCTTACGTCAACCGTACGACGGTGACGACGGCGCGT
GAGATGACGCGTCAGAGATTTCCGTTTGCATGTACCGGGAGCGTAAAGTCGTGCGGTGGT
TATGCTTCTGCTGGTTTTTTCTTCGATCCCGTCAAGAGCTGCTTCGTTAAGAGCAGAGCTT
TCTCGGGTTTGTCCGGTTCGGTTTGATCCGGTTAATATCGAGTTGAGTATTGGTATTCTGA
GAAACCGTAAAAGTTGAACCGAGAAGAGAACTAAACCTGGATCTTAACCTAGCTCCACCG
GTGGTGGACGTTTAGATTTTTTTCTTTTTCATAATTTGTATTTTACATTGCCGAAAA
TAATTAATGTTTTCTTTAG

>G31 Amino Acid Sequence (domain in AA coordinates: TBD)

MAPRQANGRSIAVSEGGGKMTMTTMRKEVHFRGVRKRPWGRYAAEIRDPGKKTRVWL
TFDTAEAAARAYDTAAREFRGSKAKTNFPLPGESTTVNDGGENDSYVNRRTVTAREMTR
QRFPFACHRERKVVGGYASAGFFDPSRAASLRAELSRVCPVRFDPVNIELSIGIRETVK
VEPRRELNLNLAPPVVDV*

>G33 (20..757)

ATTCTCCCCAACCAAAATATGACCACAGAAAAAGAGAATGTCACTACGGCCGTGGCCGT
GAAAGACGGCGGAGAAAAGAGTAAGGAAGTGAGTGACAAGGGCGTAAAGAAGAGAAAGAA
TGTAAC TAAGGCCCTGGCCGTGAATGACGGCGGAGAAAAGAGTAAGGAAGTGCGTTACAG
GGGTGTAAGGAGGAGACCATGGGGGAGATATGCTGCGGAGATCCGTGATCCGGTAAAGAA
AAAACGGGTCTGGCTCGGGTCTTCAACACGGGGGAGGAAGCCGCCAGAGCCTACGACTC
CGCTGCCATAAGGTTTCGAGGATCGAAAGCTACTACTAACTTCCCTCTAATCGGATACTA
TGGGATTTCTTCGGCGACGCCGGTGAACAACAACCTTTCCGAGACGGTGAGTGATGAAAA
TGCCAACTCCCTCTCGTTGGAGACGATGGGAATGCTTTGGCTTCTCCGGTGAACAACAC
CTTTCCGAAACGGCGCGTGATGGAACACTTCCATCGGATTGTACGACATGTTATCTCC
GGGGGTGGCTGAAGCGGTGCTGGATTTTTCTTAGATCTGCCTGAAGTTATTGCGTTGAA
AGAGGAGCTTGATCGAGTTTGTCTGACCACTTTGAGTCCATTGATATGGGGTTGACTAT
TGGTCTCTCAAACCGCGTGGAAGAGCTGAGACTTCCCTCCGCCGTGGATTGTAAGCTGCG
AATGGAACCGGATCTTGACCTCAACGCAAGTCCCTAAAGATTGATCTGATGTTGTTGTCC
TTGAATAAGTTTGTATCTTGTCTCTTCTGATTGTCTGTACTTCTATTGGTTGATTGCG
TGCTTTTGGAGGACAAAAACAACATTTTTTATGTATTAAAAAAGGTAATTGAACTATT
ATCGAAAAA

>G33 Amino Acid Sequence (domain in AA coordinates: 50-117)

MTTEKENVTAVAVKDGGEKSKEVSDKGVKKRKNVTALAVNDGGEKSKEVRYRGVRRRP
WGRYAAEIRDPVKKRVWLGSFNTGEEAARAYDSAAIRFRGSKATTNFPFLIGYYGISSAT
PVNNNLSETVSDGNANLPLVGDDGNALASPVNNTLSETARDGTLPSDCHDMLSPGVAEAV
AGFFLDLPEVIALKEELDRVCPDQFESIDMGLTIGPQTAVEEPETSSAVDCKLRMEPDLD
LNASP*

>G342 (1..723)

ATGGACGTCTACGGCATGTCTTACCAGACTTGCTTCGTATCGACGACCTTCTCGATTTC
TCCAACGACGAAATCTTCTCTTCTCTTCCACCGTCACTTCCCTCCGCCGCTTCTCCGCC
GCTTCTTCCGAAAACCCCTTTCAGCTTCTTCTTCCACCTACACTTCTCCTACTCTCCTC
ACCGACTTCACTACGATCTCTGCGTTCAGTACGACGACGCTCATCTCGAATGGTTA
TCGCGATTCTGTTGACGATTCACTTCCGATTTCAGCAATCCTTTAACCATGACCGTT
AGACCCGAGATTTCACTACCGGAAAACCTAGAAGTCGCGGATCAAGAGCACCAGCACCT
TCCGTAGCTGGAATCTGGGCTCCGATGTCTGAATCAGAGCTTTGTCACTCCGTCGCTAAA
CCTAAACCGAAGAAAGTCTACAACGCTGAATCGGTTACGGCGGATGGAGCGAGGCGGTGC
ACGCACTGTGCCTCGGAGAAAACGCCACAGTGGAGAACTGGACCGCTTGACCTAAAACA
CTTTGTAAACGCTTGTGGAGTTCTGTACAAATCAGGGAGGCTTGTACCGGAATACAGACCG
GCGTCGAGTCCGACGTTTGTATTGACTCAGCATTCGAACTCTCATCGGAAAGTTATGGAG

>G342 Amino Acid Sequence (domain in AA coordinates: 155-190)
MDVYGMSSPDLLRIDDLLDFSNDIEFSSSSTVTSSAASSAASSENPFSPSPSTYTSPTLL
TDFTHDLCVPSDAAHLEWLSRFVDDSFSDFPANPLTMTVRPEISFTGKPRRRSRAPAP
SVAGTWAPMSESELCHSVAKPKPKKYNAESVTADGARRCTHCASEKTPQWRTGPLGPKT
LCNACGVRYKSGRLVPEYRPASSPTFVLTHQHSNSHRKVMELRRQKEQQESCVRIPIPFQPO
*

>G352 (80..817)

>352 Amino Acid Sequence (domain in AA coordinates: 99-119,166-186)
MALETLSNPTATTTARPLRLRYEEMEPENLEQWAKRKRTKRQRFDHGHQNETNKNLPSE
EEYLALCLLMLARGSAVQSPPLPLPSRASPSDHRDYKCTVCGKSFSSYQALGGHKTSR
KPTNTSITSGNQELSNNSHSNSGVSVINVTVNTGNGVVSQSGKIHTCSICFKSFASGQALG
GHKRCHYDGGNGNGNGSSNSVELVAGSDVSDVDNERWSEESAIGGHRGFDLNLNLPADQV
SVTTS*

>G357 (1..615)

>G357 Amino Acid Sequence (domain in AA coordinates: 7-29)

26357 Amino Acid Sequence (Accession: E11111)
 MQNKHCKLCSKSCFCNGRALGGHMKSHLVSSQSSARKKLGDVSYSSSSSSSDGKALAYGL
 RENPRKSRFRVFNPDPESTIYNSETETETEPESGDPVKKRVRGDVSKKKKKAKSKRVFENS
 KKQKTIHSEPEPASSVSDGSPEQDLAMCLMMLSRDSRELEIKLKKPEERKPEKRHFPEL
 RRCMIDLNLPPPOEAEAVTVVSAI*

>G358 (1..855)

125

CAAGCTCTCGGTGGTCACATGAGGCGGCACAGGACAGCCGTAACCACGATTAGCCCCGTT
GCAGCCACCGCAGAAGTAAGCAGAAACAGTACAGAGGAAGAGATTGAGATCAATATAGGC
CGTTCGATGGAACAGCAGAGGAAATATCTACCGTTGGATCTTAATCTACCAGCACCAGAA
GATGATCTAAGAGAGTCAAAGTTTCAAGGGATAGTATTCTCAGCAACACCAGCGTTAATA
GATTGTCATTACTAG

>G358 Amino Acid Sequence (domain in AA coordinates: 124-135, 188-210)

MGQDEVGSDQTOI IKGKRTKRQRSSSTFVVTAATTVTSTSSSAGGSGGERAVSDEYNSAV
SSPVTTDCTQEEEDMAICLIMLARGTVLPSPDLKNSRKIHQKISSENSSFYVYECKTCNR
TFSSFQALGGHRASHKKPRTSTEEKTRLPLTQPKSSASEEQNSHFVSGSALASQASNI
INKANKVHECSICGSEFTSGQALGGHMRHRHTAVTTISPVAATAEVSERNSTEEIEINIG
RSMEQQRKYLPLDLNLPAPEDDLRESKFQGIVFSATPALIDCHY*

>G360 (1..543)

ATGTGGAACCCTAACAAAATTGAAGAATTGGAGGATGATGATGAATCTTGGGAAGTCAAA
GCCTTTGAGCAAGACACTAAAGGCAACATCTCTGGTACCACTTGGCCTCCAAGATCTTAC
ACTTGCAATTTCTGCCGCCGTGAGTTCCGTTCTGCTCAAGCCTTAGGCGGTACATGAAT
GTCCACCGCCGTGACCGCGCCTCATCTAGGGCTCATCAAGGTTCCACCGTTGCGGCTGCG
GCTAGAAGCGGCCACGGGGGGATGTTACTCAATTCTTGTGCTCCGCCGTTGCCCTACAACG
ACACTTATAATACAATCCACGGCGAGTAACATTGAAGGTTTGTCCCATTCTACCAACTG
CAAAACCCCTAGTGGCATTCTTGGTAATTCTGGTGACATGGTGAATCTTTATGTAGAAGTT
CCTCCTCGGCTTATTGAATATTCGACAGGAGATGATGAGAGCATTGGCTCGATGAAAGAA
GCGACAGGAACATCAGTGGATGAGCTTGATCTTGAACCTTCGGCTAGGGCACCATCCACCG
TGA

>G360 Amino Acid Sequence (domain in aa coordinates: 42-62)

MWNPNKIELEDDDESWEVKAFEQDTKGNISGTTWPPRSYTCNFCRREFRSAQALGGHMN
VHRRDRASSRAHQGSTVAAAARSGHGMLLNSCAPLPPTTLIIQSTASNIEGLSHFYQL
QNPSGIFGNSGDMVNLYVEVPPRLIEYSTGDDESIGSMKEATGTSVDELDELRLGLHPP
*

>G362 (195..830)

ATAAAAAACCTTCATACAATATAAAATTTCTTTAGACATACAATATATTATACTATTAC
AGATGCAATGCATCATTAGTTACAAACTATTAACTAAATATCCCCGCTCTCTCTTGC
TATATAAAGAAGATCATTTACACATCTCCTTAAGCAAATTAAACCCATCGATAAACACAT
ACGTTTACACATATATGCTATAAATCCGACAATGTCTCGTACTGGCGAAAGTTCTTCAG
GTTTCGTCTCCGACAAGACGATAAAGCTATTTCGGCTTCGAACTCATCAGCGGCAGTCGTA
CGCCGGAAATCAGACGGCGGAAAGCGTGAGCTCGTCCACAAACACGACGTCGTTAACAG
TGATGAAAAGACACGAGTGCCAATACTGCGGTAAAGAGTTTGCAAATTCTCAAGCCTTAG
GAGGTCACCAAAACGCTCACAAGAAGGAGAGGTTGAAGAAGAAGAGGCTTCAGCTTCAAG
CTCGGCGAGCCAGCATCGGCTATTATCTCACCAACCACCAACAACCCATAACGACGTCAT
TTCAGAGACAATACAAAACGCCGTCGTATTGTGCATTCTCCTCCATGCACGTGAATAATG
ATCAGATGGGTGTGTACAACGAAGATTGGTCGTGAGGTCGTGCGCAGATTAACTTCGGTA
ATAATGACACGTGCCAAGATCTTAATGAACAAAGCGGTGAGATGGGTAAGCTGTACGGTG
TTCGACCGAACATGATTCACTTCAGAGAGATCTGAGTTCTCGTTCTGATCAGATGAGAA
GTATTAACTCGCTGGATCTTCATCTAGGTTTTGCCGGAGATGCGGCATAACAAATTAAAG
AGAGATATATGATTAAAGATTATATGTACTATAGTGGCGTATTTTCATTGGGATCATGAAGG
GGAAAAAACGAGACATATAGTATTCTTGATGCAATTTGAGTTTTGTAATTTATTTAGGTT
TATGTATGTTTTTCGAAG

>G362 Amino Acid Sequence (domain in AA coordinates: 62-82)

MSINPTMSRTGESSSSGSSSDKTIKLFGEFELISGSRTPETITAEVSSSTNTTSLTVMKRH
ECQYCGKEFANSQALGGHQNAHKKERLKKRLLQLQARRASIGYYLTNHQQPITTSFORQY
KTPSYCAFSSMHVNNDQMGVYNEDWSSRSSQINFGNNDTCQDLNEQSGEMGKLYGVRPNM
IQFQRLSSRSDDQMRINSLLDLHLGFAGDAA*

>G364 (64..516)

AAGCTTGATATCGCCTCTCTCTAATCTCTCTTTCTCTCTCTATCTCTAAGAATATATAAA
GGTATGGACTACCGCAACACACATCCCTACGTCTAAGCCTACCAAGTTACAAGAACCAC
CACTAAACCTAGAACTTGTTCTCGAGCCTTCTTCCATGTCTTCTTCTTCATCTTCTTCC
ACGAACATCATCATGTGTTGGAGCAGCCTAGGGTATTCTCATGTAACCTATTGTCAAAGA
AAGTTTTACAGCTCTCAAGCTCTTGGTGGTCATCAAACGCTCATAAGCTTGAGAGAACC
TTAGCCAAGAAGAGTCGAGAACTCTTAGATCCTCAAACACTGTTGATTCTGATCAGCCT

TACCCGTTCTCCGGTCGCTTTGAGCTTTACGGCCGTGGCTACCAAGGATTTCTCGAAAGT
GGCGGTCGAGGGACTTCTCCGCCCGCCGTGTGCCGGAGAGTGGTCTTGATCAGGATCAG
GAGAAGAGTCACCTTGACTTATCCTTAAGGCTCTAAAAGAATCTTATATTTTGTAGTCT
ATATATTATCATATCAATTGTTAATCTTAAAATTGATTGTTTTACTTATTAGTCATTTCC
TATTATCTGAAAGTTTTCTTTGTAAGTTGTAAGTATGGTCCTAAATTCAAATCCAAATT
GATTTTGGAAGATGGTACCTAATGCAGTAGTTAAATAAGTTAAAAAATGAAGGATCTAT
AATTCTCT

>G364 Amino Acid Sequence (domain in AA coordinates: 54-76)
MDYQNTSLRLSLPSYKNHQLNLELVLEPSSSSSSSSSTNSSSCLEQPRVFSCNYCQRK
FYSSQALGGHQNAHKLERTLAKKSRELFSSNTVDSQPYPFSGRFELYGRGYQGFLSESG
GSRDFSARRVPESGLDQDQEKSHLDLSLRL*

>G365 (69..755)
CAATCTTTTACTTTTCATTCTCTTTATATATTCTCTCTACGTATAATATATATTACACA
GAATATACATGGAACCGTCCATCAAAGGAGATCAAGAAATGTTAAAAATCAAGAAACAAG
GTCATCAAGATCTTGAGTTGGGGTTGACCTTTTGTACGTGGAACCGCGACCTCATCAG
AGCTCAATCTCATCGATTCTTTCAAACAGCTCATCATCGACTTCTCATCATCAGCACC
AGCAAGAACAAATTGGCAGATCCGAGAGTGTCTCGTGTAAATTATTGTCAAAGAAAGTTCT
ATAGTTTACAAGCGCTAGGCGGTACCAAAACGCTCATAACGTGAGCGCACCTTAGCCA
AACGTGGACAGTATTACAAGATGACTCTCTCCTCCTTGCCCTTCTTCAGCGTTTGCCTTG
GCCACGGTTTCAGTCAGCAGATTTCGCAAGCATGGCATCGTTACCATTACATGGCTCGGTGA
ATAACAGGTCAACGTTAGGGATTCAAGCTCATTCAACGATCCATAAGCCAGCTTCTTAG
GAAGACAAACGACGAGTTTAAAGTCATGTTTTCAAACAGAGCATTACCAGAAACCGACCA
TAGGAAAGATGTTGCCGGAGAAATTTACCTTGAAGTCGCCGGAATAATAACAGTAACA
TGGTTGCTGCTAAGTTGGAGAGAATTGGACATTTCAAGAGCAACCAAGAAGATCATAATC
AGTTTAAAGAAATTGACTTGACTCTTAAAGCTATGAGCTCTGCCATCTTCTTTTAGTCTT
CATTATAACTTTTTTTATTCTCATCTTTGTTTGATATAATGATTGACGGCAGGGTGTGTT
AGAGTTTCACTAATGATCAAGTTGTACTTTTTATATATTTCATTGATACCTTGTGTGATG
AATTCAATATTTTAGGTCTGTTTTT

>G365 Amino Acid Sequence (domain in aa coordinates: 70-90)
MEPSIKGDQEMLIKIKKQGHQDLELGLTLLSRGTATSSSELNLIDSFKTSSSSTSHHQHQQE
QLADPRVFSCNYCQRKFYSSQALGGHQNAHKRERTLAKRGQYYKMTLSSLPSAFAFGHG
SVSRFASMASLPLHGSVNNRSTLGIQAHSTIHKPSFLGRQTTSLSHVFKQSIHQKPTIGK
MLPEKFHLEVAGNNNSNMVAALKERIGHFKSNQEDHNQFKKIDLTLLKL*

>G367 (1..708)
ATGGACGCTTCAATAGTTTCTCATCCACTGCTTTTCCATATCAAGATTCTCTAAACCAG
AGCATCGAAGACGAGAAAGAGACGTTTATAATTCTAGTCACGAACCTCAATCTCATCGAC
TGCATAGACGACACAACGAGTATCGTTAACGAATCTACAACATCCACAGAACAAAAGCTT
TTCTCATGCAACTATTGTCAAAGAACTTTCTATAGCTCACAAGCACTTGGTGGTCACCAA
AACGCACACAAGAGAGAGAGAACGTTGGCGAAGAGAGGACAACGTATGGCAGCGTCAGCC
TCAGCTTTTGGACATCCTTACGGTTTCTCTCCACTTCTTTCCACGGACAGTACAACAAC
CATAGGTCTTTAGGGATCCAAGCGCATTCGATAAGCCACAAGCTAAGTTCTTATAACGGG
TTTGGTGGTCACTATGGTCAGATCAACTGGTCAAGACTTCCATTTGATCAACAACAGCC
ATAGGTAAATTTCCCTCAATGGATAATTTTCATCATCATCATCATCAGATGATGATGATG
GCTCCTTCAGTAAATTCACGGTCCAATAACATCGATAGCCCAAGCAACACAGGACGGGTT
CTAGAAGGGTCACCGACTCTTGAACAATGGCACGGAGACAAAGGATTGTTGTTAAGTACA
AGTCATCATGAAGAGCAGCAGAAACTTGACTTGTCCCTCAAGCTTTGA

>G367 Amino Acid Sequence (domain in AA coordinates: 63-84)
MDASIVSSSTAFPYQDSLNSQIEDEERDVHNSSELNLIDCIDDTTISIVNESTTSTEQKL
FSCNYCQRTFYSSQALGGHQNAHKRERTLAKRGQMAASASAFGHPYGFSPLPFHGQYNN
HRSGLGIQAHSSISHKLSSYNGFGGHYQINWSRLPFDQQAIGKFPMDNFHHHHHQMmmm
APSVNSRSNNIDSPSNTGRVLEGSPTLEQWHGDKGLLLSTSHHEEQKLDLSLKL*

>G373 (1..1854)
ATGGCGATTGAAACTCAGCTTCTCTTGGCAGCGTGACGGTGTGTGATGCGGTGTCAGGTG
AATCCTCCGTCAGAAAGAGACTCTCACTTGTGGCAGCTGCGTCACTCCATGGCAGTGCCG
TGTCTCCTCCCGAATCACTCGCTTCTTCCACTGGAGAGTGGGAGTGTCCCGATTGCTCC
GGCGTTGTCGTTCCCTCCGCCGTCCGGGTACCGGAAACGCTCGACCTGAATCTTCCGGT
TCAGTTCTCGTTGCTGCGATCCGTGCGATTCAAGCTGATGAGACTTTAACCGAAGCTGAG

AAAGCCAAAAAAGGCAGAACTGATGAGTGGGGGTGGTGACGATGGTGTGCGATGAAGAA
GAGAAGAAGAAGTTAGAAATCTTTTGTCTATTTGCATTCAATTGCCAGAAAGACCTATC
ACGACACCGTGTGGGCACAATTTCTGTTTGAAATGTTTCGAGAAATGGGCAGTAGGTCAA
GGGAAGCTAACTTGATGATATGCCGAAGCAAAATTCGAGACATGTGGCAAAAAATCCT
CGCATCAACTTAGCTCTAGTTTCTGCTATTCGTTTAGCAAATGTTACCAAATGTTCTGTT
GAGGCAACTGCAGCCAAGGTTTCATCATATATCCGCAACCAAGACCGTCTGAGAAAGCA
TTTACTACCGAGCGGGCAGTAAAACTGGGAAAGCTAATGCTGCTAGCGGTAAGTTTTTT
GTGACAATACCTCGTGATCATTTTGGTCCCATACCAGCTGAGAATGATGTCACTAGAAAG
CAAGGTGTTTTGGTTGGAGAATCTTGGGAGGACAGGCAAGAGTGTAGGCAGTGGGGAGCT
CATTTCGCCCATATTGCTGGCATTGCCGGGCAATCAGCGGTTGGAGCTCAGTCTGTGGCC
CTCTCTGGAGGTTATGACGATGATGAGGATCATGGTGAATGGTTTCTCTACACAGGAAGT
GGTGGGAAGGGATCTCAGTGGAAACAAAAGAATTAACAAGAAACAGTCGTCTGACCAGGCG
TTTAAAAACATGAATGAATCTCTAAGACTTAGTTGCAAAATGGGCTATCCTGTCCGAGTT
GTCAGGTTCTTGAAGGAGAAGCGTTCTGCATATGCCCTGCTGAAGGTGTGAGATATGAT
GGGGTCTATCGAATTGAGAAGTGTGGAGTAATGTTGGAGTACAGGGTTCTTTAAGGTC
TGTCGTTACCTGTTTGTAGATGTGACAATGAGCCAGCTCCATGGACCAGTGATGAGCAT
GGCGATCGTCCAAGACCGTTGCCAATGTTCCGGAGCTTGAGACTGCTGCTGACCTGTTT
GTGAGAAAGGAGAGTCCATCATGGGATTTTCGATGAAGCTGAGGGTCTGTGGAAATGGATG
AAGTCTCCTCCTGTTAGCAGAATGGCTTTGGATCCTGAGGAGAGGAAGAATAAGAGA
GCAAAAAATACTATGAAGGCCAGACTTCTGAAAGAATTTAGTTGCCAAATCTGTCCGGAA
GTGCTGAGTCTTCCAGTGACGACGCCTTGTGCACACAACCTCTGCAAAGCATGCTTAGAA
GCGAAGTTTGTCTGGGATAACTCAACTGAGAGAGAGAAGCAATGGCGGACGTAAACTACGT
GCAAAGAAGAACATCATGACCTGCCCTTGTGTCACGACGGATCTCTCCGAGTTTCTCAA
AACCCGAGGTGAACAGAGAGATGATGGAGATAATAGAGAATTTAAGAAGAGTGAGGAA
GAGGCTGATGCATCCATTTCTGAAGAAGAAGAAGAATCCGAACCTCCAACATAAGAAG
ATTAAGATGGATAACAACCTCTGTTGGTGGTAGTGGTACAAGTCTCTCAGCTTAA
>G373 Amino Acid Sequence (domain in AA coordinates: 129-168)
MAIETQLPCDGDGVCMRQVNPSEETLTCGTCVTPWHVPCLLPESLASSTGEWECPCDS
GVVVPSPAAGTGNARPESSGSVLVAIRAIQADETLTEAEKAKKRQKLMSGGDDGVDEE
EKKKLEIFCSICIQLPERPITTPCGHNFCLKCFEKWAVGQKLTMCICRSKIPRHVAKNP
RINLALVSAIRLANVTKCSVEATAAKVHHIIRNQDRPEKAFTTERAVKTGKANAASGKFF
VTIPRDHFGPIPAENDVTRKQGVLVGESWEDRQECRQWGAHFPHIAGIAGQSAVGAQSV
LSGGYDDDEDHGEWFLYTGSGGRDLSGNKRINKKQSSDQAFKNMNESSLRLSCKMGYPVRV
VRSWKEKRSAYAPAEGVRYDGVYRIEKCWSNVGVQGSFKVCRYLFVRCDNEPAPWTSDEH
GDRPRPLPNVPELETAADLFVRKESPSWDFDEAEGRWKWMKSPVSRMALDPEERKKNKR
AKNTMKARLLKEFSCQICREVLSPVTTCAHNFCACLEAKFAGITQLRERSNGGRKLR
AKKNIMTCPCCTTDLSEFLQNPQVNREMMELIENFKKSEEEADASISEEEEESEPPKK
IKMDNNSVGGSGTSLSA*
>G396 (1..957)
ATGGGGGAAAGAGATGATGGGTTGGGTTTGTAGTCTAAGCTTGGGAAATAGTCAACAAAAA
GAACCATCTCTGAGGTTGAATCTTATGCCGTTGACAACCTTCTTCTTCTTCTTCTGTTT
CAACACATGCACAATCAGAATAACAATAGCCATCCCCAGAAGATTCAACATCTCTTGG
ACTCATCTGTTTCAATCTTCTGGGATTAAACGTACAACGTGAGAGAGAACTCCGACGCC
GGGTCAATTTCTAAGAGGTTTCAACGTGAACAGAGCTCAGTCTTCGGTGGCGGTAGTGGAC
TTGGAAGAAGAAGCCGCCGTCGTCTCGTCTCCTCAACAGCGCCGTTTCGAGTCTGAGTGGA
AATAAAAGGATCTTGCGGTGGCGAGAGGAGATGAAAACGAGGCGGAGAGAGCTTCT
TGCTCAGCGGAGGGGGAAGCGGTGGTAGCGACGATGAAGACGGCGGAAACGGCGACGGA
TCAAGGAAGAACTACGTTATCGAAGGATCAAGCTCTTGTCTCGAGGAGACTTTTAAA
GAACATAGCACTCTTAATCCGAAGCAAAAGCTGGCTCTAGCAAAACAGTTGAATCTAAGG
GCAAGACAAGTTGAAGTGTGGTTTTCAGAACCGTAGGGCAAGGACGAAGCTGAAACAAACG
GAGGTTGATTGTGAGTATTTAAAGAGATGTTGCGATAATCTGACCGAGGAGAATCGACGG
CTGCAGAAAGAAGTGTGCGAGCTGAGGGCGTTGAAGTTGTCTCCACATCTCTACATGCAC
ATGACTCCTCTACTACTCTCACCATGTGCCCTTCTTGGCAACGTGTCTCCTCTCTGCC
GCCACTGTGACCGTGTCTCTTCCACTACTACTCTCTACGTTGGTGGGGCGGCCAAGT
CCACAGCGATTAACTCCTTGGACTGCTATTTCTCTCCAGCAAAATCAGGTCGCTAG
>G396 Amino Acid Sequence (domain in AA coordinates: 159-220)
MGERDDGLGLSLSLGNSQQKEPSLRLNLMPLTSSSSSFQHMHNQNNNSHPQKIHNISW

>G431 (1..1149)

>G431 Amino Acid Sequence (domain in AA coordinates: 286-335)
MESGSNSTSCPMFAFGDNSDGPMCPMMMMPPIMTSHQHGHGHDHQHQQQQEHGDAYAYQSHH
QQSSSLFQLSLAPPQGTKNKVASSSSPSSCAPAYSLMEIHNEIVAGGINPCSSSFSSAS
VKAKIMAHPHYHRLLAAYVNCQKVGAPPEVVARLEEACSSAAAAAASMGPTGCLGEDPGL
DQFMEAYCEMLVKYEQELSKPFKEAMVFLQRVECQFKSLSLSSPSSFSGYGETAIDRNNN
GSSEEEVDMMNEFVDPQADLRELKQGLLRKYSGYLGLSLKQEFMKKRKKGKLPKEARQQL
DWWSRHYKWPYPSEFQQLALAEASTGLDQKQINNWFINQRKRHWKPSSEDMQFVVM DATHPH
HYFMDNVLDNPFPMDHISSTML*

ATGGAGATGGGTTTCCAACACTCGGGTCCGGGTCTATGTTCCGGGTGTCAGGCAGAGTCCGGGTGGT
TCCTCCACTGAGTCATCCTCTTTTCAGTGGAGGGTCTATGTTTGGCCAGAAAGATCTACTTC
GAGGACGGTGGTGGTGGATCCGGGTCTTCTCTCCTCAGGTGGTCTGTTCAACACAGACAGCTGTCT
CGTGGAGGCGGGTCCGGTTCAGTCGGGTTCAGATACCAAGGTGCCAAGTGGAAAGGTTGTGGG
ATGGATCTAACCAATGCAAAGGTTATTACTCGAGACACCGAGTTTGTGGAGTGCAGCTCT
AAAAACCTAAAGTCACTGTGGCTGGTATCGAACAGAGGTTTTGTCAACAGTGCAGCAGG
TTTCATCAGCTTCCGGAATTTGACCTAGAGAAAAGGAGTTGCCGCAGGAGACTCGCTGGT
CATAATGAGCGACGAAGGAAGCCACAGCCTGCGTCTCTCTCTGTGTTAGCTTCTCGTTAC
GGGAGGATCGCACCTTCGCTTTACGAAAATGGTGATGCTGGAATGAATGGAAGCTTTCTT
GGGAACCAAGAGATAGGATGGCCAAGTTCAAGAACATTGGATACAAGAGTGATGAGGCGG
CGAGTGTCTGTCACCGTTCATGGCAGATCAATCCAATGAATGTATTTAGTCAAGGTTTCAGTT
CGGTGGAGGAGGCAACAGCTTCTCATCTCCAGAGATTATGGACACTAAACTAGAGAGCTAC
AAGGGAATTTGGCGAAGTCAAACCTGTGCTCTCTCTCTTCTGTCAAATCCACATCAACCACAT
GACAACAACAACAACAACAACAACAACAACAACAACAACAATACATGGCGAGCTTCT
TCAGGTTTTTGGCCCGATGACGGTTTCAATAGGCTCAACCACCCTGCACCTAGCCAGCAT
CAGTATCTGAACCCGCTTGGGTATTCAAGGACATGATAATGATATGTCTCCTGTTTTG
AATTTAGGTCGATACACCGAGCCAGATAATTGTGCAGATAAGTAGTGGCAGGCAATGGGT
GAGTTTCAGTTATCTGATCACCATCATCAAAGTAGGAGACAGTACATGGAAGATGAGAAC
ACAAGGGCTTATGACTCTTCTCTCACCATAACCACTGGTCTCTCTGA

129

FHQLPEFDLEKRSCRRRLAGHNERRRKQPASLSVLASRYGRIAPSLYENG DAGMNGSFL
GNQEIGWPSSRTLDTRVMRRPVSSPSWQINPMNVFSQGSVGGGTSFSSPEIMDTKLESY
KGIGDSNCALSLLSNP HQPHDNNNNNNNNNNNTWRASSGFGPMTVTMAQPPPAPSQH
QYLNPPWVFKDNDNDMS PVLNLGRYTEPDNCQISSGTAMGEFELSDHHHQSRQYMEDEN
TRAYDSSSHHTNWSL*

>G546 (1..588)

atgactcgaccgtcaagattacttgagacggcgccaccaccacaaccgtcggaggag
atgatcgacgcggaatccgacatggtggtgatcttgtcggctcttcttgcgctcttatac
tgcgttgctggtctcgccgcgctcgtagatgcgcttggtccggcggtttacagccgga
ggagattcgccgtcaccgaacaaaggcttgaaaaagaaagctcttcagtctcttccaaga
tccactttcaccgcgcggaatcaacctccggcgccgcgctgaagaggagactcgacg
gaatgtgctatttgctcactgacttcgccgacggtgaagaaataagagtgtcttctctt
tgtggtcattcttccacgtggagtgtattgacaaatggctagtttctaggtcttcttgt
ccttcttgtcgcaggattcttacgcgggtgagatgtgaccggtgtggtcatgttctacg
gcggagatgaaagatcaagctcatcgatcaacatcaccaacactcttctactaccatt
cctacgtttcttctcttaa

>G546 Amino Acid Sequence (domain in AA coordinates:114-155)

MTRPSRLLETAAPPQPSEEMIAESDMVVILSALLCALICVAGLAAVVRCALRRFTAG
GDSPSPNKGLKKKALQSLPRSTFTAESTSGAAAEEDSTECAICLTDFADGEEIRVLPL
CGHSFHVECIDKWLVSRS SPCRRILTPVRCDRCGHASTAEMKDQAH RHQHQSSTTI
PTFLP*

>G551 (1..708)

ATGGAGTGGTCAACAACGAGCAACGTAGAAAACGTGAGAGTAGCTTTCATGCCACCGCCA
TGGCCGGAGTCTAGTTCTTTAACTCGCTCCACAGCTTCAACTTTGATCCTTACGCAGGA
AATTCATATACGCCTGGCGATACACAAACCGACCGGTTATCTCTGTACCGGAATCAGAA
AAGATCATGAATGCGTACCGATTTCCGAACAACAACATGAGATGATAAAAAAGAAGAGA
CTAACGAGTGGACAATTAGCTTCACTTGAGCGAAGTTTTCAAGAAGAGATCAAATTAGAT
TCAGACAGGAAGGTGAAGCTGTGAGAGAGCTCGGTCTGCAGCCACGTGAGATAGCAGTT
TGGTTCCAAAACCGCCGTGCACGGTGGAAGGCGAAGCAGCTTGAGCAGTTGTACGACTCG
CTTAGACAAGAGTACGACGTCGTTTCTAGGGAGAAACAAATGTTACACGATGAGGTGAAG
AAGCTGAGAGCTTTACTAAGAGACCAGGGTTTGATCAAGAAGCAAATCTCTGCCGGGACC
ATCAAAGTTTCCGGTGAGGAAGACACGGTGAGATTTTCATCGGTGGTGGTAGCTCATCCA
AGAACGGAGAATATGAACGCAAAATCAAATCACCGAGGGAATCAAGTTTACGGTCAATAC
AACAAATCCGATGCTGGTTGCTTCTCTGGCTGGCCGTCATACCCCTGA

>G551 Amino Acid Sequence (conserved domain in AA coordinates:73-133)

MEWSTTSNVENVRVAFMPPWPPESSSFNSLHSFNFDYPAGNSYTPGDTQTGPVISVPESE
KIMNAYRFPNNNNEMIKKRLTSGQLASLERSFQEEIKLSDRKVKLSRELGLQPRQIAV
WFQNRRLRWKAKQLEQLYDSL RQEYDVVSREKQMLHDEVKKLRALLRDQGLIKKQISAGT
IKVSGEEDTVEISSVVVAHPR TENMNANQITGGNQVYGYNNPMLVASSGWPSYP*

>G578 (1..978)

ATGCATAGTTTGAATGAAACAGTAATTCCTGATGTTGATTACATGCAGTCTGATAGAGGG
CATATGCATGCTGCTGCCTCTGATTCCAGTGATCGATCAAAGGATAAGTTGGATCAAAAG
ACCTTCGTAGGCTTGCTCAAAATCGTGAGGCAGCAAGAAAAGCAGATTGAGGAAGAAG
GCGTATGTTTACGAGCTGGAAGATAGTCGATTAAAGCTGACTCAAGTTGAGCAGGAGCTG
CAAAGAGCAAGACAGCAGGGAGTTTTCATCTCAAGTTGAGGAGACCAAGCTCATTCTACT
GGTGGCAATGTTGGGGCTTTGGCATTGATGCAGAACACTCACGATGGCTTGAAGAAAAG
AACAGGCAAATGAACGAGCTGAGATCTGCCCTGAATGCTCATGCAGGTGATACTGAGCTC
CGGATAATTGTGGATGGAGTGTGCTCACTATGAGGAGCTTTTCAGGATTAAGAGCAAT
GCATCTAAGAATGATGTCTTCCACTTGTATCTGGAATGTGGAACCAACAGCTGAGCGA
TGTTTCTTGTGGCTTGGCGGGTTCCCGTCATCCGAACCTTCTCAAGCTTCTTGCGAATCAG
CTAGAGCCCATGACAGAACGACAGGTAATGGGCATCAATAGCTTGACGACAGCTCGCAG
CAGGCAGAAGATGCTTTATCTCAAGGGATGGAGAGTTTACAGCAATCCCTAGCTGATACT
TTATCCAGTGGAACCTTGGTTCCAGTTCATCGGATAATGTCGCGAGCTACATGGGTGAG
ATGGCCATGGCAATGGGCAAGTTAGGCACCCCTCGAAGGATTATACGCCAGGCTGATAAC
TTGAGGCTGCAAACTACAACAGATGCTTCGAGTATTAACAACACGTCAGTCAGCTCGT
GCTCTTCTTGCTATACAGATTATTCATCTCGATTACGTGCTCTTAGTTCTTGTGGCTT
GCCCCGCCAAGAGAGTGA

>G578 Amino Acid Sequence (domain in AA coordinates 36-96)
MHSINETVIPVDVDMQSDRGHMHAAASDSSDRSKDKLDQKTLRRLAQNREAARKSRLRKK
AYVQQLSEDSRLKLTQVEQLQRRARQQGVFISSSGDQAHSTGGNGGALAFDAEHSRWLEEK
NRQMNELRSALNAHAGDTELRIIVDGVMAHYEELFRIKSNASKNDVFHLLSGMWKTPAER
CFLWLGGFPSSSELLKLLANQLEPMTERRQVMGINSLQQTQQAEADALSQGMESLQQSLADT
LSSGTLGSSSSDNVASYMGQMAMAMGKLGTLLEGFIRQADNLRQLQQLRVLVLTTRQSAR
ALLAIHDYSSRLRALSSLWLARPRE*

>G596 (168..1121)

TAATTTCTCTACTTCAGATTTTTTTCTCCTTAGATTAATTTAATTGAGTTATTGTACATC
CCTCAAGCTAAGATTCTGGTTTTGTGAGTTGAGTGGATGAGAAGAGGAGAGATTAACATA
ATTAGGGTTTCAATTGTTTACTTTTTGTTTGCTTTTTATATCAAGTAATGGATCAGGTCT
CTCGCTCTCTTCTCCACCTTTTCTCTCAAGAGATCTCCATCTTCACCACACCATCAAT
TCCAGCATCAGCAGCAGCAGCAACAGAATCACGGCCACGATATAGACCAGCACCAGAA
TCGGTGGGCTAAAACGTGACCGAGATGCTGATATCGATCCCAACGAGCACTCTTCAGCCG
GAAAAGATCAAAAGTACTCCTGGCTCCGGTGGAGAAAAGCGCGCGGAGGAGGAGAGATA
ATCACATCAGAGAAGGCCACGTGGCAGACCAGCGGGATCTAAGAACAACCAAAACCGC
CAATCATCATCACTCGAGACAGCGCAAACGCTCTCAAATCTCATGTCTGGAAGTAGCAA
ACGGATGTGACGTCTGGAAGGTGTCACCGTCTTCGCTCGCCGTCGCCAACGTGGCATCT
GCGTTTTGAGCGGAAACGGCGCGGTTACCAACGTTACCATAAGACAACAGCTTCAGTAC
CTGGTGGTGGCTCATCTGCTGTTAACTTACACGGACGTTTTGAGATTCTTTCTCTCTCGG
GATCATTCTCTCTCTCCGGCTCCACAGCTGCGTCAGGTCTAACGATTTACTTAGCCG
GTGGTCAGGGACAGGTTGTTGGAGGAAGCGTGGTGGTCCACTCATGGCTTCAGGACCTG
TAGTGATTATGGCAGCTTCGTTTGGAAAACGCTGCGTATGAGAGACTGCCGTTGGAGGAAG
ACGATCAAGAAGAGCAAACAGCTGGAGCGGTTGCTAATAATATCGATGGAAACGCAACAA
TGGGTGGTGGAAACGCAAAACGCAAACTCAGACGCAGCAGCAACAGCAACACAGTTGATGC
AAGATCCGACGTCGTTTATACAAGGGTTGCCCTCCGAATCTTATGAATCTGTTCAATTGC
CAGCTGAAGCTTATTGGGGAACCTCCGAGACCATCTTCTAAATCGCGAAGAAAAACAAG
TTAGATACGTTTCGTTGTTTTTAATTTATAATCTCTCTCTGTCAAGTTTAAATTTCTTT
TTCTTCTCTTTGTTTTCTAAAGATAATTGTAGTCTTTGACGAAGATTCGTGGTACGTAT
GAATCGAAGAGAATCGTTTTGGTTCATGGGATTGCTCGATCTATTAGGTTTGAGAGGGGGT
TTGTGTTTTGCGTTGACTAGCAGATTATAAAATGTTGATTTTCGAGTTTATTTTCAT
GTGTTGGTGATAAA

>G596 Amino Acid Sequence (domain in AA coordinates: 89-96)

MDQVSRSRLPPPFLSRDLHLHPHHQFQHQQQQQQNHGHDIDQHRIGGLKRDRDADIDPNE
HSSAGKDQSTPGSGGESGGGGGDNHITRRPRGRPAGSKNPKPPIIITRDSANALKSHV
MEVANGCDVMESVTVFARRRQRGICVLSGNGAVTNVTIRQPASVPGGSSVNVNLHGRFEI
LSLSGSFLPPPAPPAASGLTIYLAGGQGVVGGSSVVGPLMASGPVIVMAASFNGAAYERL
PLEEDDQEEQTAGAVANNIDGNATMGGGTQTQTQTQQQQQQQLMQDPTSFIOGLPPNLMN
SVQLPAEAYWGTTPRPSF*

>G617 (59..1141)

CAGATCTGTTCTTTACACCAAATTGAGTACTGAAGATCTTGTTGAGTGAATTAAAGAGAT
GAGATCAGGAGAATGTGATGAAGAGGAGATTCAAGCAAAGCAAGAAAGAGATCAAAATCA
AAATCATCAAGTAAACTTAAACCACATGTTGCAACAACAACAGCCGAGTTCGGTATCATC
TTCAAGGCAATGGACTTCAGCTTTTAGGAATCCAAGAATCGTTCGAGTCTCAAGAACATT
CGGTGGCAAAGACAGACACAGCAAAGTATGTACAGTCCGTGGTCTTCGAGACCGGAGGAT
AAGGTTGTCCGTACCTACAGCTATTCAACTCTACGACCTTCAAGATCGATTAGGGCTGAG
TCAGCCAAAGCAAAGTCATTGATTGGTTACTCGAAGCAGCAAAGATGACGTAGACAAGCT
ACCTCCTTACAAATCCCACATGGATTAAACAGATGTATCCAATCTCATCTTCGAAA
CTCCGGGTTTGGAGAATCTCCATCTTCAACTACATCAACAACGTTTCCAGGAACCAATCT
CGGGTTCTTGGAAAATTGGGATCTTGGTGGTCTTCAAGAACAAGAGCAAGATTAACCGA
TACAACTACGACCCAAAGAGAAAGTTTGATCTTGATAAAGGAAAATGGATCAAAAACGA
CGAGAATAGTAATCAAGATCATCAAGGGTTTAACACCAATCATCAACAACAATTTCTCT
GACCAATCCGTACAACAACACTTCAGCTTATTACAACCTTGGACATCTTCAACAATCGTT
AGACCAATCTGGTAATAACGTTACTGTGCGAATATCTAATGTTGCTGCTAATAATAACAA
TAATCTCAATTTGCATCTCTCTCTCGTCTGCCGGAGATGGATCTCAGCTTTTTTTCGG
TCCTACTCTCCGGCAATGAGCTCTCTATTCCCGACATACCTTTCGTTTCTTGGAGCTTC
TCATCATCATCATGTCTGTCGATGGAGCCGGTCATCTTCAGCTCTTTAGCTCGAATTCAAA

TACCGCATCGCAGCAACACATGATGCCGGGTAATACGAGTTTGATTAGACCATTTCATCA
TTTGATGAGCTCGAATCATGATACGATCATCATAGTAGCGATAATGAATCAGATTCTTG
AATGATTTTATATATCTACACTATACATTGAAAATGTTATATGTATACGTATTCTTCTAT
ATTTTGATATATATGCGTATTGTTGGATTGGTTTATGTATCT

>G617 Amino Acid Sequence (domain in AA coordinates: 64-118)
MRSGECDDEEIQAKQERDQNHQVNLNHLQQQPSSVSSSRQWTSAFRNPRIVRVSR
FGGKDRHSKVCTVRGLRDRIRLSVPTAIQLYDLQDRLGLSQPSKVIDWLLEAAKDDVDK
LPPLQFPHGFNQMYPNLIFGNSGFGESPSSTTSTTFPGTNLGFLENWDLGGSSRTRARLT
DTTTTQRESFDLDKQKWIKNDENSNQDHQGFNTNHQQQFPLTNPYNNTSAYYNLGHLLQOS
LDQSGNNVTVAISNVAANNNNNLNLHPPSSSAGDGSQLFPGTPPAMSSLFPTYPSFLGA
SHHHHVVDGAGHLQLFSSNSNTASQQHMPGNTSLIRPFHMLSSNHDTDHSSDNESDS
*

>G620 (40..666)
GAATTGAACTTGGACCAGCACAGCAACAACCCCAACCCCAATGACCAGCTCAGTCATAGTA
GCCGGCGCCGGTGACAAGAACAATGGTATCGTGGTCCAGCAGCAACCACCATGTGTGGCT
CGTGAGCAAGACCAATACATGCCAATCGCAAACGTCATAAGAATCATGCGTAAACCTTA
CCGTCTCAGCCAAAATCTCTGACGACGCCAAAGAAACGATTCAAGAATGTGTCTCCGAG
TACATCAGCTTCGTGACCGGTGAAGCCAACGAGCGTTGCCAACGTGAGCAACGTAAGACC
ATAACTGCTGAAGATATCTTTGGGCTATGAGCAAGCTTGGGTTTCGATAACTACGTGGAC
CCCCTCACCGTGTTCATTAAACCGGTACCGTGAGATAGAGACCGATCGTGGTTCTGCACTT
AGAGGTGAGCCACCGTCTGTTGAGACAAACCTATGGAGGAAATGGTATTGGGTTTCACGGC
CCATCTCATGGCTACCTCCTCCGGGTCCTTATGGTTATGGTATGTTGGACCAATCCATG
GTTATGGGAGGTGGTCCGTAACAAAACGGGTCGTCGGGTCAAGATGAATCCAGTGTT
GGTGGTGGCTCTCTGCTCTTCCATTAACGGAATGCCGGCTTTTGACCATTATGGTCAGTAT
AAGTGAAGAAGGAGTTATTCTTCTTATATCTATTCAAACATGTGTTTCGATAGAT
ATTTTATTTTATGTCTTATCAATAACATTTCTATATAATGTTGCTTCTTTAAGGAAAAG
TGTTGTATGTCAATACTTTATGAGAACTGATTTATATATGCAAAAT

>G620 Amino Acid Sequence (domain in AA coordinates: 20-118)
MTSSVIVAGAGDKNNGIVVQQPPCVAREQDQYMPIANVIRIMRKTLP SHAKISDDAKET
IQECVSEYISFVTGEANERCQREQRTITAEDILWAMSKLGFNDYVDPLTVFINRYREIE
TDRGSALRGEPPSLRQTYGGNGIGFPHGPHGLPPPGPYGYGMLDQSMVMGGGRYYQNGSS
GQDESSVGGSSSSSINGMPAFDHYGQYK*

>G625 (151..1137)
AATCGACCATTCACAACGATGACATTCAAACACTCTTCAGTTTCCCTTCCTTCTTGATT
GTCTCTCCACTATTTTCTCAATTTCTTTAATCTCTCTCTTCTCTCTCTACTTCTCT
TCCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
AACCATCTGGAAGATAATAACCAAAACCTAACCATAATAATCTCAATCCGATTCCACC
ACCGACTCATCAACTTCTCTCCGCTCAACGCAACGCAAGGCAAGGTGGTCCGGACAAC
TCCAAGTTCCGTTACCGTGGCGTTCGACAAAGAAGCTGGGGCAATGGGTCCGGCAGATC
CGAGAGCCACGTAAGCGCACTCGCAAGTGGCTTGGTACTTTTCGCAACCGCCGAAGACGCC
GCACGTGCCTACGACCGGCTGCCGTTTACCTATACGGGTCACGTGCTCAGCTCAACTTA
ACCCCTTCGTCTCTCTTCTCTCCGTCTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCT
CCTTCCACCTCCTCTCTCTCCACTCAAACCTAAGACCTCTCTCTCTCTCTCTCTCTCTCT
GCCACCGTAGGAGGAGGACCAACTTTGGTCCGTACGGTATCCCTTTTAAACAACAACATC
TTCCTTAATGGTGGGACCTCTATGTTATGCCCTAGTTATGGTTTTTTCCCTCAACAACAA
CAACAACAAATCAGATGGTCCAGATGGGACAATCCAACACCAACAGTATCAGAATCTT
CATTCTAATACTAACAATAACAAGATTTCTGACATCGAGCTCACTGATGTTCCGGTAACT
AATTCGACTTCGTTTCATCATGAGGTGGCGTTAGGGCAGGAACAAGGAGGAAGTGGGTGT
AATAATAATAGTTCGATGGAGGATTTGAACCTCTAGCTGGTTCGGTGGGTTCGAGTCTA
TCAATAACTCATCCACCGCGGTTGGTTGATCCGGTATGTTCTATGGGTCTGGATCCGGGT
TATATGGTTGGAGATGGATCTTCGACCATTGGCCCTTTTGGAGGAGAAGAAGAATATAGT
CATAATTGGGGGAGTATTTGGGATTTTATGATCCCATCTGGGGGAATTCTATTAATTT
GTTTTTGTGGAAGATCATATTATATACGATGAGCATCCCTAAGGTCCGTCAAGAGCATTG
GAGATTCTGTTGAGAGGAATCAAAGAGATTGCATTCTATGAGGAGCTCTGCATGCAAA
ATTTTGGAGGATTTTTTACTACCTATAGAGATAAATAAGAGGGTATTTTTATTTATTTT
TTGAAGATTTTTATTTTCAAGGAATTCGTAAGAGATTACGGTTCCAATAAAGTATGTA
TATGTGGAAGAGAATCGGAGGAGATGGTGGAAAGTTGTATGGGAATTTTATTGGTTCAAC

ACTCCTCCTCCTCCAACGAATTGTAGCTATAGGTTGTTTGGATTGTGATCTCACAAGCAAT
TCTCCTGCCTCAATCCCTCAAGACAAGCAACCGATGGATACTTGTGGAGCTGCCAAGTGT
CAAGAACCCATCACTCCAACCTCAATGAGTGAGCAGAAGAAGCAACAAACATCAAGAAGT
CGAACTAAAGTGCAAATGCAAGGCATTGCGGTTGGTCTGCGGTTGATTAAACACTGTTG
AAATCTTACGATGAAGTGAATGATGAGCTTGAGGAGATGTTTGAGATTCAAGGACAGCTT
CTTGCCCGAGACAAATGGATCGTTGTCTTCACTGATGATGAAGGAGATATGATGCTTGCT
GGTGATGATCCGTGGAATGAGTTTTCGAAGATGGCAAAGAAGATATTTATATATTCGAGC
GATGAGGTTAAGAAAATGACAACGAAACTGAAGATTTCTTCGTCGTTAGAGAATGAGGAA
TATGGTAATGAATCATTTCGAAAATCGTAGTAGGGGGTGAGAGTTTTAGCTGTTAATTAAG
GTTAATTCGGCGACGTCGTTTTAGTGCCTAAGTGTCTAAAGACTTTTTTTTTTAGTCTGTG
TATATAAAGTCTTGTCTCTTTTTTCATGTCAATTTTTCAAGTTGGCGATTTAATATTTG
GTTTTGGGACAGTGGTTGATGGGCGGTTTTACATTTTTTATGTGTATGTACTTGTTC
AAACCATTCAATTTTCAA

>G716 Amino Acid Sequence (domain in AA coordinates: 24-355)
MASVEGDDDFGSSSSRSYQDQLYTELWKVCAGPLVEVPRAQERVVFYFPQGHMEQLVASTN
QGINSEEIIPVFDLPPKILCRVLDVTLKAEHETDEVAQITLQPEEDQSEPTSLDPPIVGP
TKQEFHSFVKILTASDSTHGGFSVLKHKATECLPSLDMTQATPTQELVTRDLHGFWEWRF
KHIFRGQPRRHLLTTGWSTFVSSKRLVAGDAFVFLRGENGDLRVGVRRRLARHQSTMPTSV
ISSQSMHLGVLATASHAVRTTIFVVFYKPRISQFIVGVNKYMEAIAKHGFSLGTRFRMR
EGEESPERIFTGTIVGSGDLSSQWPASKWRSLLQVQWDEPTTVQRPDKVSPWEIEPFLATS
PISTPAQQPQSKCKRSRPIEPSVKTPAPPSFLYSLPQSQDSINASLKLFDPSLERISGG
YSSNNSFKPETPPPTNCSYRLFGLTNSNPAPIQDKQPMDCGAACQEPITPTSMS
EQKKQQTSSRSRTKVQMGIAGRAVDLTLLKSYDELIDELEEMFEIQGQLLARDKWIVVF
TDDEGDMMLAGDDPWNEFCMAKKIFIYSSDEVKMTTKLKISSSLENEEYGNESFENRS
RG*

>G725 (46..1122)
CCTCTTTTCAGAGAGAAAGAGAGTCAGAGAGAGAGAGAGAGAGAATGTTCCATGCTAAG
AAACCTTCAAGTATGAATGGTTCATATGAGAACAGAGCTATGTGCGTTCAAGGCGATTCA
GGCCTTGTCTCACCACCGACCCTAAACCGCGTTTGGCGTTGGACCGTCGAACCTCACGAG
CGTTTTGTGGACGCGCTCGCTCAGCTCGGCGCCCCGACAAAGCGACCCCAAGACGATT
ATGAGAGTTATGGTGTGAAGGGTCTTACTCTTTACCACCTAAAGAGCCATCTTCAGAAA
TTCAGGCTTGGAAGCAGCCGCACAAGGAGTACGGAGATCACTCCACAAAGGAAGGTTCA
AGAGCTTCGCCATGGATATTAGCGCAACGTAGCTTCTTCTTCTGGCATGATGAGTCGC
AACATGAATGAGATGCAAATGGAAGTGCAGAGAAGGTTGCATGAACAGCTAGAGGTGCAA
AGACATCTGCAACTGAGGATTGAAGCACAAGGAAAGTACATGCAATCTATCTTGGAGAGA
GCTTGCCAAACCTAGCCGGTGAGAACATGGCAGCCGCCACCGCAGCAGCCGCCCTCGGA
GGAGGATACAAGGGTAATCTGGGAAGTTCGAGTCTTTCAGCAGCGGTGGGCCCCACCTCT
CATCCTCTTAGTTTCCCGCGGTTTCAAGACCTAAACATCTATGGAACACAAACCGACCAA
GTCCTCGACCATCAAACTTCCATCATCAAAACATAGAGAACCAATTTACGGGTAAACAAT
GCTGCAGACACCAACATTTACTTGGGGAAGAAGCGACCTAATCTAATTTTGGTAACGAT
GTAAGGAAAGGACTATTGATGTGGTCTGATCAAGATCAGATCTTTCGCAAAACCAATCG
ATCGATGATGAGCATAGAATTAGATACAGATGGCTACACATGTCTCCACGGATTTGGAT
TCTTTGTGCGAGATCTACGAAAGGAAATCAGGTTTATCAGGTGATGAAGGGAATAATGGT
GGGAAATTACTGGAAGGCCATCGCCTAGGAGATCACCATTGAGTCCTATGATGAACCTT
AATGGTGGATTAATACAAGGAAGAACTCGCCATTGGGTGATACAATTTATTAATTTTT
ATCTATGAGTGATGCATGGGAATGTAAGAACGAGATATATATGTTTTGTCAATTGTGAGTT
TGACGTAGGGTTTAGAGAAAA

>G725 Amino Acid Sequence (domain in AA coordinates: 39-87)
MFHAKKPSSMNGSYENRAMCVQDGLVLTDPKPLRWTVELHERFVDAVAQLGGPDKA
TPKTIMRVMGVKGLTLYHLKSHLQKFRLLGKQPHKEYGDHSTKEGSRASAMDIQRNVASS
GMMSRNMNEMQMEVQRRLEHQLVQRHLQLRIEAGQKYMQSILERACQTLAGENMAAATA
AAAVGGGYKGNLSSSLAAVGGPPHPLSPFPFQDLNIYGNTTDQVLDHNFHHQNIENH
FTGNNAADTNIYLGKRPNPNFGNDVRKGLLMWSDQDHDLSANQSIDDEHRIQIQMATHV
STDLDLSLEIYERKSGLSGDEGNNGKLLERPSRRSPLSPMMNPNGGLIQGRNSPFG*
>G727 (43..1977)
CTTCTTCTCCTTCTCTGATCGTTCTGTTTCTGGACGAGAGAGATGGTAAATCCGGGTCAC
GGAAGAGGACCCGATTCCGGGTACTGCTGCTGGTGGGTCAAACCTCCGACCCGTTTCTGCG

AATCTTCGAGTTCTTGTCGTTGATGATGATCCAACCTTGCTCATGATCTTAGAGAGGATG
CTTATGACTTGTCTCTACAGAGAGCAGAGAGCGCATTGTCTCTGCTTCGGAAGAACAAG
AATGGTTTTGATATTGTCATTAGTGATGTTTCATATGCCTGACATGGATGGTTTTCAAGCTC
CTTGAACACGTTGGTTTAGAGATGGATTTACCTGTTATCAATCTGAATGTTTTGAAACCT
TTGGTTATAGTGATGCTCTCGGATGATTCTGAAGAGCGTTGTGTTGAAAGGAGTGACTCAC
GGTGAGTTGATTACCTCATCAAACCGGTACGTATTGAGGCTTTGAAGAATATATGGCAA
CATGTGGTGCAGGAAGAAGCGTAACGAGTGGAATGTTTTCTGAACATTCTGGAGGAAGTATT
GAAGATACTGGCGGTGACAGGGACAGGCAGCAGCAGCATAGGGAGGATGCTGATAACAAC
TCGTCTTCAGTTAATGAAGGGAACGGGAGGAGCTCGAGGAAGCGGAAGGAAGAGGAAGTA
GATGATCAAGGGGATGATAAGGAAGACTCATCGAGTTTAAAGAAACCACGCGTGGTTTGG
TCTGTTGAATGTCATCAGCAGTTTGTGCTGCTGTGAATCAGCTAGGCGTTGACAGTGAG
TTAAAACTTGCTTGCTTATGCAATTTGTGTGTGTCGATTGTAACATTGTGGAATTCAG
AAGTATCGGATATATCTGAGACGGCTTGAGGAGTATCGCAACACCAAGGAAATATGAAC
CATTCGTTTATGACTGGTCAAGATCAGAGTTTGGACCTCTTCTTCGTTGAATGGATTT
GATCTTCAATCTTTAGCTGTTACTGGTCAGCTCCCTCCTCAGAGCCTTGACAGCTTCAA
GCAGCTGGTCTTGCCCGGCTTACACTCGCTAAACCAGGGATGTCGGTTTTCTCCCTTGTA
GATCAGAGAAGCATCTTCACTTTGAAAACCCAAAAATAAGATTTGGAGACGGACATGGT
CAGACGATGAACAATGGAAATTTGCTTCATGGTGTCCCAACGGGTAGTCACATGCGTCTG
CGTCTCGACAGAATGTTTCAGAGCAGCGGAATGATGTTGCCAGTAGCAGACCAGCTACCT
CGAGGAGGACCATCGATGCTACCATCCCTCGGGCAACAGCCGATATTGTCAAGCAGCGTT
TCAAGAAGAAGCGATCTCACTGGTGCCTGGCGGTTAGAAACAGTATCCCCGAGACCAAC
AGCAGAGTGTTACCAACTACTCACTCGGTCTTCAATAACTTCCCCGCGGATCTACCTCGC
AGCAGCTTCCCGTTGGCAAGTGCCCCAGGGATTTAGTTCCAGTATCAGTTTCTTACCAA
GAAGAGGTCAACAGCTCGGATGCAAAAGGAGGTTTCATCAGCTGCTACTGCTGGATTTGGT
AACCCAAGCTACGACATATTTAACGATTTTCCGCAGCACCAACAGCACAAACAAGAACATC
AGCAATAAACTAAACGATTGGGATCTGCGGAATATGGGATTGGTCTTCAGTTCCAATCAG
GACGCAGCAACTGCAACCGCAACCGCAGCATTTCCTCACTTCGGAAGCATACTCTTCGTCT
TCTACGCAGAGAAAAAGACGGGAAACGGACGCAACAGTTGTGGGTGAGCATGGGCAGAAC
CTGCAGTCACCGAGCCGGAATCTGTATCATCTGAACCACGTTTTTATGGACGGTGGTTCA
GTCAGAGTGAAGTCAGAAAGAGTGGCGGAGACAGTGACTTGTCTCCAGCAAATACATTG
TTTCAGGAGCAGTATAATCAAGAAGATCTGATGAGCGCATTTCTCAAACAGGTTTGATTA
TTACTCGAATACAGTGCATCTAAAAAC

>G727 Amino Acid Sequence (domain in AA coordinates: 226-269)
MVNPGHGRGPDSTAGGSNSDPFANLRLVLDVDDPTCLMILRMLMTCLYREQRAHCL
CFGRKNGFDIVISDVHMPDMDGFKLLEHVGLEMDLPVINLNLKPLVIVMSADDSKSVV
LKGVTGAVDYLKIPVRIEALKNIWQHVVRRKRNENNVSEHSGGSIEDTGGDRDRQQHR
EDADNNSSSVNEGNRRSRKRKEEVEDDQGDDESSSLKPRVWVSVLHQFVAAVNQ
LGVDSELKTCLLMHLVCSIGNIVEFQKYRIYLRRLGGVSVQHQNMMNHSFMTQDQSFGL
SSLNGFDLQSLAVTGQLPPQSLAQLQAAGLGRPTLAKPGMSVSPVLDQRSIFNFENPKIR
FGDGHGQTMNNGNLLHGVPTGSHMRLRPGQNVQSSGMLPVADQLPRGGPSMLPSLGQOP
ILSSSVRRSDLTGALAVRNSIPETNSRVLPTTHSVFNNFPADLPRSSFPLASAPGISVP
VSVSYQEEVNSSDAKGGSSAATAGFGNPSYDIFNDFPOHQHKNISNKLNDWDLRNMGL
VFSSNQDAATATATAAFSTSEAYSSSSTQRKRRETDATVVGEHGQNLQSPSRNLYHLNHV
FMDGGSVRVKSERVAETVTCPPANTLFHEQYNQEDLMSAFLKQV*

>G740 (25..924)

CTTCTTCAACTTTTTTTTTTAACGATGGCTTCAGAGGATCAATCGGCGGCGAGATCTACC
GGGAAGGTGAAGTGGTTCAACGCTTCTAAAGGCTATGGTTTCATTACTCCTGACGATGGC
AGCGTAGAGCTTTTGGTTCATCAATCTTCAATTGTCTCCGAAGGTTACCGGAGTTTAAAC
GTCCGGGATGCGGTTGAGTTGCTATTACTCAGGGAAGCGACGGTAAGACTAAAGCCGTC
AATGTTACTGCTCCTGGTGGTGGTTCTCTCAAGAAGGAGAATAACTCTCGTGGTAACGGT
GCTAGGCGCGGCGGCGGTGGAAGCGGTTGCTACAATTGCGGTGAGTTAGGTCATATCTCT
AAAGATTGTGGTATTGGTGGCGGCGGCGGAGGTGGTGAACGTAGATCTAGAGGAGGAGAA
GGTTGTTACAATTGTGGTGATACTGGTCACTTCGCTAGGGATTGTACTTCAGCTGGAAAC
GGTGACCAACGTGGAGCCACCAAGGTGGAAACGATGGTTGCTACACTTGGCGGTGATGTT
GGTCAGTGGCTAGGGATTGTACTCAGAAATCAGTTGGAAACGGAGACCAACGTGGAGCG
GTCAAAGGTGGAAACGATGGTTGCTACACTTGTGGTGATGTTGGTCACTTTGCTAGGGAT
TGTAATCAGAAGGTTGCTGCCGGAACGTCAGAAGCGGTGGTGGTGGTAGTGGAACCTTGT

TATTCATGCGGTGGAGTTGGTCACATTGCAAGAGATTGTGCGACTAAGAGACAGCCTTCT
CGTGGGTGTTACCAAGTGTGGTGGTCTGGTCACCTGGCTCGTGATTGTGACCAGAGAGGA
AGCGGTGGAGGAGGTAATGATAATGCGTGCTACAAGTGTGGTAAGGAAGGTCACCTTTGCA
AGGGAATGTTCTTCTGTAGCTTAATCGATTTCCTAATCAACAAAACAAAAACAAGAAT
GAAATTGAATCGAGTTATATAGTTTGGTATATATTACTCTTCGTTTTTCATTTATCTTTTT
TTTTGTTGTTGATGGGAATGAAATTGCCTGGTCCCTTTTGGTGTGTTTTTGGAGCTTTTATT
ATTATACAGAGTGATCCCTTTTTTGTATATACTATTACAAGTTTTTAGCTTTATTTGATA
TGGATGCTCTCTCCTTTTTCTTCTATCTGTTTCTGGAAATTTTGACCTCATCATATTACTT
ATGTCATCCAAA

>G740 Amino Acid Sequence (domain in AA coordinates: 24-42, 232-268)

MASEDQSAARSTGKVNWFNASKGYFITPDDGSVELFVHQSSIVSEGYRSLTVGDAVEFA
ITQGSDBGKTKAVNVTPGGGSLKKENNSRNGARRGGGSGCYNCGELGHISKDCGIGGG
GGGERRSRGEGCYNCGDTGHFARDCTSAAGNDQRGATKGGNDGCTCGDVGHVARDCT
QKSVNGDQRGAVKGGNDGCTCGDVGHFARDCTQKVAAGNVRSGGGSGTCYSCGGVGH
IARDCATKRQPSRGYQCGGSGHLARDCDQRGSGGGGNDNACYKCKEGHFAECSSVA*
>G770 (119..1069)

CCTTCCCTCTATATAAGGAAGTTCATTTCAATTTGGAGAGGACACGCTGACAAGCTGACTCT
AGCAGATCTGGTACCGTCGACGGTCTTGGATTGGAGTAACTAAAGATCATATAAAAT
GGAACAAGGAGATCATCAGCAGCATAAGAAAGAAGAAGACTTTGCCACCGGGTTTCAG
ATTTTCATCCGACGGATGAGGAGCTAATCTCATATTACTTGGTTAATAAGATTGCCGATCA
AACTTCACCGGGAAGCAATCGCTGACGTTGATCTTAACAAGTCCGAGCCATGGGAGCT
TCCTGAGAAGGCGAAAATGGGAGGAAAAGAATGGTACTTTTTTAGCCTCCGGGACCGGAA
GTACCCGACGGGAGTGAGGACGAATAGGGCGACGAATACAGGATATTGGAAAACACAGG
AAAAGACAAAGAGATATTCAATAGCACAACTCGGAGTTGGTTCGGGATGAAGAAGACTTT
GGTCTTTTACAGAGGACGAGCTCCTCGTGGGGAGAAGACTTGTGTTGGGTGATGCATGAGTA
TCGACTTCACTCCAAGTCTCATATAGAACCTCCAAGCAAGACGAGTGGGTAGTGTGTAG
AGTGTTCAGAAAACAGAAGCAACCAAGAAATACATAAGCACCAGTAGCAGCAGCACAAG
TCATCACCACAACAACACACAAGAGCCTCAATACTATCAACCAACAACAATAATCTTAA
TTACTCATCAGACCTCCTTCAACTCCACCGCATCTACAACCACACCCGAGCCTCAATAT
TAACCAATCCCTCATGGCAAACGCCGTTACCTAGCTGAGCTCTCAAGAGTCTTCCGTGC
CTCTACAAGCACCACCATGGACTCTTCTCATCAGCAGCTAATGAAGTACACCCACATGCC
TGCTTCAGGGCTCAACCTCAACCTTGGCGGTGCAGTGGTCCAGCCGCTCCTGTTGTGTC
TCTTGAGGATGTTGCGCGGTTAGTGCTTCGTACAATGGCGAAAACGGGTTTGGAAATGT
GGAGATGAGCCAGTGCATGGACTTGGATGGATACTGGCCATCTTATTGATTGGTAATTTGT
CAGTTTAAAGTTATGGTTTTTATATTGTTTCCATTTACTTGTGTTGTAACGATTTTGGTT
GTTCTTGCGAACGCTCTAGACAGGCCTCGTACCGGATCCTCTAGCTAGAGCTTTCGTTTCG
TATCATCGGTTTC

>G770 Amino Acid Sequence (domain in AA coordinates: 19-162)

MEQGDHQHKKKEEALPPGFRFHPTDEELISYYLVNKIADQNFTGKAIADVLDLNKSEPWE
LPEKAKMGGKEWYFFSLRDRKYPTGVRTNRATNTGYWKTGKDKEIFNSTTSELVGMKKT
LVFYRGRAPRGEKTCWVMHEYRLHSKSSYRTSKQDEWVVCRVFKTEATKKYISTSSST
SHHHNNHTRASILSTNNNNPNYSSDLLQLPPLQPHPSLNINQSLMANAVHLAELSRVFR
ASTSTTMDSSHQQLMNYTHMPVSGNLNLGALVQPPPVVSLDVDAAVSASYNGENGFN
VEMSQCMDLDGYWPSY*

>G858 (99..869)

CATAATCTCTTCTCTATATCTCTTCTTCTTCTTTTACCCTGTTTTTTTTTTCATTC
CACAGAGCCCAGGTTGATTGATTTTGTATTTCAGAGATATGGGGAGAGGAAGGATTGAGA
TTAAGAAGATTGAGAATATCAACAGTCGTCAAGTCACTTTCTCTAAGAGACGAAACGGTT
TGATCAAGAAGGCTAAAGAGCTTTTCGATTCTCTGTGACGCCGAGGTTGCTCTTATCATCT
TCTCCAGCACCGGCAAGATTTACGATTTCTCCAGCGTCTGTATGGAGCAAATTTCTTCTA
GATATGGATACACTACTGCGTCCACTGAGCATAAACAACAAGAGAACACCAACTTCTAA
TTTGTGCTTCACATGGAAATGAAGCTGTGTTGCGAAATGATGATTCTATGAAGGGGGAAC
TTGAAAGATTACAGCTTGCAATTGAGAGACTTAAGGGTAAGGAGCTTGAAGGTATGAGTT
TCCCGGATCTTATTTCTTTGAAAACAGTTGAACGAGAGCTTGCATAGTGTCAAGGATC
AAAAGACACAAATCTGTCTCAACCAGATTGAGAGATCCAGGATACAGGAGAAAAAGCAT
TGGAAGAAAACCAATCTTGCGCAAACAGGTTGAGATGTTGGGGAGAGGTTTCAGGACCAA
AAGTGTGTAATGAAAGGCCTCAAGATTCTAGCCCAGAAGCCGATCCCGAGAGCTTTCAT

CAGAAGAGGATGAGAATGACAACGAGGAGCACCATTCCGACACTTCCTTGCACTTGGGGT
TGTCGTCGACGGGGTATTGCACAAAGAGAAAGAGCCGAGATCGAACTGGTCTGCGATA
ACTCTGGGAGTCAAGTGGCTTCTGATTGATGGAATCGATTATTTTTCTAATTCTGGTTGT
TTAGGGGTCTCTATGTGTCTTCTGTTTCTGGCTGTTCTTTTGCTTTATTTTCATCTCAAG
TAGAGTTTTCTTAATGTTTAGGTGGAACATTTTCCATAATCAAGAAGGGATTGATCAA
TCAATAACATTAGATTTTCTTAGTTAAAGACTTAAAGTTGCCACACACCACCATATG
TGATTATGATGAATTTACATTTTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>G858 Amino Acid Sequence (domain in AA coordinates: 2-57)

MGRGRIEIKKIENINSRQVTFSKRRNGLIKKAKELSILCDAEVALIIFSSTGKIYDFSSV
CMEQILSRGYTTASTEHKQOREHQLLICASHGNEAVLRNDDSMKGELERLQLAIERLKG
KELEGMSFPDLISLENQLNESLHSVVDQKTQILLNQIERSRIQEKKALEENQILRKQVEM
LGRGSGPKVLNERPDSSPEADPESSSSEEDENDNEEHSDTSLQLGLSSTGYCTKRKKP
KIELVCDNSGSQVAD*

>G865 (282..920)

ATCCCCACTTGTGTTCATCACCAGCCAGCTCCATGTCCTAGTCACTCCACAGATTCC
CTATCATCATCAATTTCGTTTCAAACCTTAGTTTCCTTTCAAAGTCTTGACATATATACACA
CACACCTATTATTCTCTTGGTGTGTTTGTGTGTACATATACGTGTGAGTACATACTTTG
TTGTAAAGTGGATCGGAGGTATGGAAAGGACCGGTTCCACCGGAAACATCGGCGGCGG
CGGATGATAATTCGTCTTGAACGAGACTGATGTCACCGCCATGGTCTCCGCTCTCAGCC
GTGTCATAGAGAATCCGACAGACCCGCGGTCAAACAAGAGCTTGATAAATCGGATCAAC
ATCAACCAGACCAAGATCAACCAAGAAGAAGACACTATAGAGGCGTAAGGCAGAGACCAT
GGGGTAAATGGGCGGCAGAAATCCGCGATCCAAAGAAAGCAGCCCGTGTCTGGCTCGGGA
CTTTCGAGACGGCAGAGGAAGCTGCTTTAGCCTATGACCGAGCTGCCCTCAAATTCAAAG
GCACCAAGGCTAAACTGAACTTCCCTGAACGGGTCCAAGGCCCTACTACCACCAACCA
TTTCTCATGCACCAAGAGGAGTTAGTGAATCCATGAACTCACCTCCTCCTCGACCTGGTC
CACCTTCAACTACTACTACTTCTGTTGCAATGACTTATAACCAGGACATACTTCAATACG
CTCAGTTGCTTACGAGTAACAATGAGGTGATTTATCATACTACAGTCTGACTCTCTTCA
GTCAACCTTTTTCAACGCCTTCTTCATCTTCTTCTCCCAACAGACGCAGCAACAGC
AGCTACAACAACAACAACAGCAGCGTGAAGAAGAAGAGAAGAAATTATGGTTACAATTATT
ATAACTACCCAAGAGAATAATCTAATTTATTATTGTTGGTTCGAATCAGTTTTATAAATAGC
TATCATAGTTTTCATTTTGGTTTCCGTAACTTTGTTGCATGGAAAATATGAATGAACGA
GGGATGTTGTTAACAATTTGTTTGTGTTTTCGTAAATGTTAGTTGTAATTTGGATTGCTGA
AGTTTGATTTTCTGAGCATAAATCATTTGACGGTCAAAAAAAAAAAAAA

>G865 Amino Acid Sequence (domain in AA coordinates: 36-103)

MVSALSRVIENPTDPVVKQELDKSDQHQPDPQDQPRRRHYRGVRQRPWGKWAIEIRDPKKA
ARVWLGTFFETAEEAALAYDRAALKFKGTKAKLNFPERVQGPTTTTTISHAPRGVSESMNS
PPPRPGPPSTTTTWSMPTYNQDILQYAQLLTSNNEVDLSYTTSTLFSQPFSTPSSSSSSS
QQTQQQQQLQQQQQQREEEKKNYGYNYNYPRE*

>G872 (59..646)

CCGGAACAGAAATCCAATTCACCAACCGAATCGAACCGAACCGGAGTTTTTATCCAAT
GGTGAAGCAAGCGATGAAGGAAGAGGAGAAGAAGAGAAACACGGCGATGCAGTCAAAGTA
CAAAGGAGTGAGGAAGAGGAAATGGGGAAATGGGTATCGGAGATCAGACTTCCACACAG
CAGAGAACGAATTTGGTTAGGCTCTTACGACACTCCCGAGAAGGCGGCGGTGCTTTCTGA
CGCCGCTCAATTTTGTCTCCGCGGCGGCGATGCTAATTTCAATTTCCCTAATAATCCACC
GTCGATCTCCGTAGAAAAGTCGTTGACGCTCCCGAGATTCAAGGAAGCTGCTGCTAGATT
CGCTAACACATTTCCAAGACATTGTCAAGGGAGAAGAAGAAATCGGGTTTTAGTACCCCGATC
CGAGATCCGACAGAGTCTCCTTCTACATCTGCATCTGTTGCTACATCGACGGTGGATTA
TGATTTTCTGTTTTCGATTGCTTCCGATGAATTTCCGGTTTGATTCTTCTCCGACGA
CTTCTCTGGCTTCTCCGGTGGTGATCGATTACAGAGATTTTACCCATCGAAGATTACGG
AGGAGAGAGTTTATTAGATGAATCTTTGATTCTTTGGGATTTTGAATTTCCCAAACATAA
TATTTTTTTAGAGCGAACTGTGAGATTTTCTTGGAGTCATGGAGAAATCTGGAGATTTT
TTGTAACACGGAGCTCCAATGACCCGGGAATTTCTTTTCGTTTCGGATCCGAATTTGATGT
GGATCATATTACACCTATATTTTTTTCATTTTTTGTGTAAGAAAAATCGGATAAGAT
TCTAGTAATAAATGTTAAAAAGTCCATTTTCAATAAAAAAAAAAAAAAAAAAAAAA

>G872 Amino Acid Sequence (domain in AA coordinates: 18-85)

MVKQAMKEEEKKRNAMQSKYKGVKRKWKWVSEIRLPHSRERIWLGSYDTPEKAARAF
DAAQFCLRGGDANFNFPNNPPSISVEKSLTPPEIQEAAARFANTFQDIVKGEESGLVPG

SEIRPESPSTSASVATSTVDYDFSFLDLLPMNFGFDSFSDDFSGFSGGDRFTEILPIEDY
GGESLLDESLILWDF*

>G904 (1..1005)

atggaatctctcatcaatcccagccatggcggaggaaactacgattctcactcttcttct
ctcgatagtcctcaaaccagcgtactagtcattctcattctcctcatgactcttctc
atctccggtttccatttgcctcctcctcgcgtgtctcaatcgctgtagccaccgctccggt
ctccctctttcatcttctcctcctcgcgtcgcaaccgtaacttccgattcccagcattctct
ggacatcgagtcctcccgaacagaacgggtcctccggtgcttgattcgcttccgattttc
aaattctcctcgcgtcactcgccgatctagctccatgaattccggagattgcccgtttgt
ttgtcgaaattcgaaaccggaggatcagctccgtcttctcctctctgttgacgctttt
cacgccgattgtatcgatatctggctagctcttaaccagacttgctcctctctgctgctct
cctctcttgcgttcagaatctgactctcatgaagtctctcgccgtcgctcggtcaaacaac
ggcggaggagaaaacagcttccgtctcgaaatcggtatccatcagccgtcgctcgtaacaac
ccgattccagaatccgttgagcagcatcgaacttactcaatcggttcggttcgattacata
gtagacgagctagattcagaaatctcagagtcaaatttcaaccgtggaaaacaggaagac
gcgactacaacaactgccacagcaacggcggttacgactaatccgacgtcggttgagct
agtttagcggcgatataaggaacgatgggtctagaagctggctcaaggattacgttgac
agactctcagcaggtatctcgcgtgcaatgtcggttagaagctctggtagattttt
actgggagtagtcggtgagtgaggaattgacggtgatggatttagaagcgaatcatgcc
ggagaagagataaagtgaagctttccggtggctctcaggggtgtga

>G904 Amino Acid Sequence (domain in AA coordinates: 117-158)
MESLINPSHGGGNYDSHSSSLDSLKPSVLVIIILILLMTLLISVSICFLRLNRCSHRSV
LPLSSSSSVATVTSDSRFSGHRVSPETERSSVLDLPIFKSSVTRRSSMNSGDCAVC
LSKFEPEDQLRLPLCCHAFHADCIDILVSNQTCPLCRSPLFASESILMKSLAVVGSNN
GGGENSFRLEIGSISRRRQTPIPESVEQHRTYSIGSFYIVDDVDSEISESNFNRGKQED
ATTTTATATAVTNPTSFEASLAADIGNDGSRSLWKDYVDRLSRGISSRAMSFRSSGRFF
TGSSRRSEELTVMDLEANHAGEEISELFRWLSGV*

>G910 (1..1071)

ATGTTATGTATAATAATAATTGAGAATATGGAAGAGTATGTGAGTTTTGTAAAGCGTAT
AGAGCAGTGTTTTATTGTATAGCTGATACAGCAAATCTTTGTTTAAATGTGATGCAAAG
GTTCAATTCAGCTAATCTACTCTCGGGACGGCATTACGTACGTTTTATGTGATTCTGGT
AAGAATCAGCCTTGTGTGTCGGATGTTTTGACCATAAAATGTTCTTTGCCATGGATGT
AATGATAAGTTTCATGGTGGTGGCTCTTCTGAGCATCGTAGAAGGGATTTGAGGTGTTAT
ACGGGTTGTCTCTCTGCTAAAGATTTCGCGGTTATGTGGGGTTTTCGAGTTATGGATGAC
GATGATGATGTTTCGTTAGAGCAATCTTTTGAATGGTTAAACCTAAGGTGCAAAGAGAA
GGTGGTTTTATCTTGAACAGATTCTTGAATTGGAGAAGGTTTCAGCTCAGGGAAGAGAA
GGTAGTTCTTCTTGAACAGAGGATCCATCTCCATTGGAGCTTCTTAAGAAACCC
GAAGAACAGTTAATCGATCTTCCGACAGACCGGAAAGAGCTGGTTGTTGATTTTTCACAC
TTGTCTCATCTTCCACACTTGGTGATTCTTTTGGGAATGCAAAAGTCCATACAATAAG
ACAATCAGTTGTGGCATCAAAATATACAAGACATTGGAGTATGTGAAGATACAATCTGC
AGTGACGATGACTTCCAAATACCTGACATTGATCTCACTTTCCGGAACTTTGAAGAGCAA
TTTGGAGCTGATCCTGAGCCAATTGCAGATAGTAACAACGTGTTCTTTGTTTCTTCCCTT
GACAAATCACATGAGATGAAGACATTTTCTTCTTCAATAATCCCATATTTGCACCT
AAACCAGCTTTCATCAACTATCTCATTCTCAAGCAGTGAAACCGATAACCCTTATAGTCAC
TCAGAGGAAGTAATCTCATTTTGTCCCTCCCTCTTAACAATACACGTCAAAGGTCATC
ACAAGGCTCAAGGAGAAGAAGAGAGCAAGAGTGGAGGAGAAAAAGCTTAA

>G910 Amino Acid Sequence (domain in AA coordinates: 14-37, 77-103)
MLCIIIIENMERCVEFCKAYRAVVYCIADTANLCLTCDAKVHSANSLSGRHLRTVLCDSG
KNQPCVVRFCDFHKLCHGKNDKFHGGSSSEHRRDLRCYTGCPKDFAVMWGFRVMD
DDVSLQSFMRVQREGGFLEQILELEKVLREENGSSSLTERGDPSPLEPKP
EEQLIDLPTGKELVDFSHLSSSSTLGDSPFECKSPYNKNNQLWHQNIQDIGVCEDTIC
SDDDFQIPDILTFRNFEQFGADPEPIADSNVFFVSSLDKSHMKTFSSSFNNPIFAP
KPASSTISFSSSETDNPYSHSEVISFCPSLSNNTRQKVITRLKEKKRARVEEKKA*
>G912 (20..694)

CATCTTATCCAAAGAAAAATGAATCCATTTTACTCTACATTCCCAGACTCGTTTCTCTC
AATCTCCGATCATAGATCTCCGGTTTCAGACAGTAGTGAGTTTACCAAAGTTAGCTTC
AAGTTGTCCAAAGAAACGAGCTGGGAGGAAGAAGTTTCGTGAGACACGTCATCCGATTTA

CAGAGGAGTTCGTCAGAGGAATCTCGGTAAATGGGTTTGTGAAGTTAGAGAGCCTAATAA
GAAATCTAGGATTTGGTTAGGTACTTTTCCGACGGTTGAAATGGCTGCTCGTGCTCATGA
TGTTGCTGCTTTAGCTCTTCGTGGTTCGCTCTGCTTGTCTCAATTCGCTGATTCTGCTTG
GCGGCTTCGTATTCTGAGACTACTTGTCTTAAGGAGATTACAGAAAGCTGCGTCTGAAGC
TGCAATGGCGTTTCAGAAATGAGACTACGACGGAGGGATCTAAACTGCGGCGGAGGCAGA
GGAGGCGGCAGGGGAGGGGGTGAAGGAGGGGAGAGGAGGGGCGGAGGAGCAGAATGGTGG
TGTGTTTTATATGGATGATGAGGCGCTTTTGGGGATGCCCAACTTTTTTTGAGAAATATGGC
GGAGGGGATGCTTTTGC CGCCGCCGGAAGTTGGCTGGAATCATAACGACTTTGACGGAGT
GGGTGACGTGTCACTCTGGAGTTTACGAGTAATTTTTTGGCTCTTTTTCTGGATAATA
AGTT

>G912 Amino Acid Sequence (domain in AA coordinates:51-118)
MNPFYSTFPDSFLSISDHRSPVSDSSECSPKLASSCPKRAGRKKFRETRHPIYRGVQR
NSGKWVCEVREPNNKSRILWLTFTVEMAARAHDAALALRGRSACLNFAADSAWRLRIPE
TTCPEIKQAASEAAMAFQNETTEGSKTAAEAEEAAGEGVREGERRAEQNGGVFYMD
EALLGMPNFFENMAEGMLLPPEVGVNHNDFDGVGDVSLWSFDE*

>G920 (114..1154)
AAAAATCTATTTTCTTCTCTTCCACTATATTACAACATTTCTTCATTCTCAAATCATC
ATACTAAAAACCTAAAAAAGTTACATATTCATTGTATCTTTGTGAGAAAAAATGGATT
CGAATAGTAACAACACGAAATCCATAAAGAGAAAAAGTTGTGACCAACTTGTGGAAGGCT
ATGAATTCGCTACTCAGCTTCAGCTTCTCCTTTCTCATCAACACTCTAACCAGTACCACA
TCGATGAGACCCGTCTTGTTTTCCGGGTCGGGTTTCCGGTGGTCCAGATCCCGTTG
ATGAGCTCATGTCTAAGATCTTGGGATCTTCCATAAACTATATCGGTTCTTGATTCTT
TTGATCCCGTCGCCGTCTCTGTCCCATCGCCGTCGAGGGTTCATGGAATGCTTCATGTG
GGGATGATTCGGCGACTCCGGTGAGTTGCAACGGTGGAGATTCCGGTGAGAGTAAGAAGA
AGAGATTAGGGGTTGGTAAGGGTAAAGAGGATGCTACACTAGAAAGACGAGATCACATA
CAAGGATCGTGAAGCTAAAGTTCTGAAGACAGATATGCTTGGAGGAAATATGGACAAA
AGGAGATTCTTAATACCACATTCCCAAGAAGTTACTTTAGATGCACACACAAGCCAACGC
AAGGATGCAAAGCAACAAAGCAAGTTCAGAAACAGGATCAAGATTCTGAGATGTTCCAAA
TCACATACATTGGCTACCACACATGCACTGCCAATGACCAAACGCACGCGAAGACCGAGC
CTTTTGATCAAGAAATCATTATGGATTCCGAAAAGACATTGGCTGCTAGCACTGCTCAGA
ACCATGTCAATGCTATGGTGCAAGAGCAAGAGAACAACACCAGCAGTGTGACAGCAATAG
ACGAGGATGGTTAAGGAGGAACAAAATAACAATGGTGATCAGAGTAAAGATTATTATG
AGGGCTCTTCGACAGGTGAGGACTTGTCTATTGGTTTGGCAAGAGACGATGATGTTTGATG
ATCATCAAAATCACTACTATTGTGGTGAAACCAGTACTACTTCTCATCAATTGGTTTCA
TCGACAACGATGATCAGTTTTCTCCTTCTTCGACTCATATTGTGCTGATTATGAAAGAA
CAAGTGCTATGTGAACATCCAAATCTGGAATGATGAATCAGCACTAGGTCTTCTCTTGA
GTATGTCTAGTTTAAATGTAATATTTTGTGTATGTTTGATAAAAAACACCATATATACTT
CTCTTTTACACCAAAAAAAAAAAAAAAAAAAAAA

>G920 Amino Acid Sequence (domain in AA coordinates: 152-211)
MDSNSNNTKSIKRKVVQDLVEGYEFATQLQLLLSHQHSNQYHIDETRLVSGSGSVSGGPD
PVDELMSKILGSFHKTIISVLDSFDPVAVSVPIAVEGSWNASCGDDSATPVSCNGGDSGES
KKKRLGVGKGRGCTYTRKTRSHRIVEAKSSEDRYAWRKYQKEILNTTFPRSYFRCTHK
PTQGCKATKQVQKQDQDSEMFQITYIGYHTCTANDQTHAKTEPFDQEIIMDSEKTLAAS
AQNHVNAMVQEQENNTSSVTAIDAGMVKEEQNNNGDQSKDYEGSSTGEDLSLVWQETMM
FDDHQNHYYCGETSTTSHQFGFIDNDDQFSSFFDSYCADYERTSAM*

>G939 (9..1565)
CAGATTCTATGGATATGTATAACAACAATATAGGGATGTTCCGGAGTTTAGTTTGTAGCT
CGGCGCCTCCATTTACAGAGGGACATATGTGTTCTGATTTCGCATACGGCTTTGTGCGATG
ATCTGAGTAGTGATGAGGAAATGGAATAGAGGAGCTTGAGAAGAAGATCTGGAGAGACA
AGCAGCGTTTAAAGCGGCTCAAGGAAATGGCGAAGAACGGTCTAGGAACAAGATTGTTGT
TGAAGCAGCAACATGATGATTTTCCAGAGCACTCTAGTAAGAGAACCATGTACAAGGCAC
AAGATGGGATCTTGAAGTACATGTGCAAGACAATGGAGCGATATAAAGCTCAAGGTTTGT
TTTATGGGATTGTGTAGAGAATGGGAAACGGTAGCGGGATCTTCTGATAATCTCCGTG
AATGGTGGAAGACAAAGTGAGGTTTGTATAGGAACGGCCAGCTGCTATAATCAAGCACC
AAAGGGATATCAATCTTCTGATGGAAGTGATTACAGGGTCTGAGGTTGGGGATTCTACCG
CACAGAAGTTGCTTGAGCTTCAAGATACTACTCTTGAGCTCTGTTATCGGCTCTGTTTC
CTCACTGCAACCCCTCTCAGAGGCGGTTTCCGTTGGAGAAAGCGGTGACACCGCCATGGT

GGCCAACGGGGAAAGAAGATTGGTGGGATCAACTGTCTTTACCCGTTGATTTTCGAGGTG
TTCCGCCACCTTACAAGAAGCCTCATGATCTCAAGAAGCTGTGGAAAATTGGTGTTTGA
TTGGTGTAAATCAGACATATGGCTTCTGACATTAGCAACATACCCAATCTCGTGAGACGGT
CTAGAAGTTTGCAGGAGAAAATGACGTCAAGAGAAGGCGCTTTATGGCTCGCTGCTCTTT
ACCGAGAAAAGGCTATTGTTGATCAAATAGCCATGTCTAGAGAAAACAACAACACTTCTA
ACTTTCTTGTTCCTGCAACCGGTGGAGACCCAGATGTTTTGTTTCTGAATCTACAGACT
ATGATGTTGAACTGATTGGTGGCACTCATCGGACCAATCAGCAGTATCCTGAATTTGAAA
ACAATACTAACTGTGTTTACAAGAGAAAGTTTGAAGAAGATTTTGGGATGCCAATGCATC
CAACACTCCTAACATGTGAGAACAGTCTCTGTCTTATAGCCAACCATATGGGATTTCT
TTGACAGGAACTTAAGAGAGAATCACCAAATGACTTGTCTTATAAAGTCACTTCTCTTCT
ACCAACCAACTAAACCCTATGGTATGACGGGTTTAAATGGTTCCTTGTCCGGATTATAACG
GGATGCAGCAGCAGGTTTCAAGACCAGTTTAAATCATCCCAACGATCTCTACA
GACCAAAGCTCCACAAGAGGCAACGATGACTTGGTTGAGGATTTGAATCCTTCTCTCTT
CGACGCTGAATCAGAATCTTGGTTTAGTCTTACCTACTGACTTCAATGGAGGTGAGGAAA
CAGTAGGAACAGAGAACAACTCTGCATAATCAAGGGCAAGAGTTGCCACATCTTGGATT
AGTAAAGAAAGCTTCAAGTCTTCTTTTATGTTTTCTAGTCTTTATAGCTTTGTCTCTT
GCTTATCTCTCATTAAACACAGTCTTCTGATCTCTCCATTTATAGCCCATGTAGCAATG
GAGAAAGATTAGGTTTTCATAATAAGTTAATAACCAAATTCAAA

>G939 Amino Acid Sequence (domain in AA coordinates: 97-106)
MDMYNNIGMFRSLVCSAPPFTEGHMCSDSHTALCDDLSSDEEMEIEELEKKIWRDKQR
LKRLEMAKNGLGTRLLLLKQHQDDFPEHSSKRTMYKAQDGIKYSKTMERYKAQGFVYG
IVLENGKTVAGSSDNLREWWKDKVRFDRNGPAIIKHQDINLSDGSDSGSEVGDSTAQK
LLELQDPTTLGALLSALFPHCNPPQRRFPLEKGVTPPWPTGKEDWWDQLSLPVDVFRGVPP
PYKKPHDLKKLWKIGVLIGVIRHMASDISNIPNLVRRSRSLQEKMTSREGALWLAALYRE
KAIVDQIAMSRENNNTSNFLVPATGGDPDVLFPSTDYDVELIGGTHRTNQQYPEFENNY
NCVYKRKFEEFDGMPMHPHTLLTCENSLCPYSQPHMGFLDRNLRENHQMTCPIKVTSTFYQP
TKPYGMTGLMVPDYPDNGMQQVQSFQDQFNHPNDLYRPKAPQGRNDDLVEDLNPSSTL
NQNLGLVLPDFTNGGEETVGTENNHLNQGEELPTSWIQ*

>G963 (1..897)

ATGAGTTTGCCTCCAGGATTCAGGTTTCATCCCACTGATGAAGAAGCTGGTGGCTTACTAT
CTTGATAGGAAGGTCAACGGCCAAGCCATTGAGCTCGAGATCATCCGAGAAGTTGATCTT
TATAAATGCGAGCCATGGGACTTGCCTGAAAAGTCAATTTTGGCCGGGAAACGACATGGAA
TGGTACTTTTACAGCACAAAGGGATAAGAAGTATCCAAATGGCTCTAGGACGAACCGTGCG
ACCCGAGCGGGTTACTGGAAGGCCACGGGGAAAGATCGTACAGTAGAATCAAAGAAGATG
AAGATGGGAATGAAGAAGACACTGGTTTATTATAGAGGAAGGGCTCCTCATGGCCTTCGT
ACTAATTGGGTCTATGCATGAATATCGTCTCACGCACGCTCCTTCTCTCTCTTGAAGGAG
TCGTATGCATTGTGCCGAGTGTTTAAGAAGAACATACAAATTCCAAAGAGAAAAGGGGAA
GAAGAAGAAGCAGAAGAAGAGAGCACTAGTGTAGGAAAAGAAGAGGAAGAAGAAAAGGAG
AAGAAGTGGAGAAAATGTGATGGTAATTATATTGAAGACGAGAGCTTGAAAAGAGCATCC
GCGGAGACATCTTCATCAGAGCTAACTCAAGGGGTCTTTTATAGACGAAGCAAACAGCTCA
TCCATATTTGCTCTTCATTTCTCATCTTCTCTCTGGACGATCATGATCATCTTTTCTCA
AACTATTCTCATCAGCTTCCATATCATCTCTCTTCAACTCCAAGATTTCCCTCAACTT
TCTATGAACGAAGCAGAGATTATGTCAATCCAACAAGACTTTCAATGCAGAGACTCTATG
AACGGGACACTTGACGAAATCTTCTCTTCTTCCGCCACTTTCCCGCTTCCCTTTGA

>G963 Amino Acid Sequence (domain in AA coordinates: TBD)
MSLPPGFRFHPTDEELVAYYLDKRVNGQAIIELEIPEVDLYKCEPWLPEKSFLPGNDME
WYFYSTRDKKYPNGSRNTRATRAGYWKATGKDRTVESKKMKMGKTLVYYRGRAPHGLR
TNWVMHEYRLTHAPSSSLKESYALCRVFKNIQIPKRKGESEEBEESTSVGKEEEEEKE
KKWRKCDGNYIEDESLKRASAEISSSELTQGVLLDEANSSSIFALHFSSSLDDHDLHLS
NYSHQLPYHPPLQLQDFPQLSMNEAEIMSIQQDFQCRDSMNGTLDEIFSSSATFPASL*

>G979 (60..1352)

CCTCTGAGGAATCAAATCACTCACACTCCAAAAAATCTAAACTTTCTCAGAGTTTAA
TGAAGAAGCGCTTAACCACTTCCACTTGTCTTCTTCTCCATCTTCTCTGTTTCTTCTT
CTACTACTACTTCTCTCTTATTCAGTCGGAGGCTCCAAGGCCTAAACGAGCCAAAAGGG
CTAAGAAATCTTCTCTTCTGGTGATAAATCTCATAACCCGACAAGCCCTGCTTCTACCC
GACGAGCTCTATCTACAGAGGAGTCACTAGACATAGATGGACTGGGAGATTGAGGCTC
ATCTTTGGGACAAAAGCTCTTGGGAATTCGATTCAGAACAAGGAAGGCAACAAGTTTATC

141

AGAGATACGGTTTTTGAAGCTGATAAGGAAGATAAAGCCAGACGTGTTTCATCCCCGGGATC
CTCAGCGGATCCTACAACGCGCCTTTCTTTGTCACGAGGTTTAGAGAAGTTCTGTTTCAT
TACTCATCTCTGTTTGACATGTGTGACACGAATCTAACACGGGAAGATCCAATGAGGGTT
ATGTTTGAGAAAGAGTTCTATGGGCGGGAGATCATGAACGTGGTGGCGTGTGAGGGGACG
GAGAGAGTGGAGAGGCCAGAGAGTTATAAGCAGTGGCAGGCGAGGGCGATGAGAGCCGGG
TTTAGACAGATTCCGCTGGAGAAGGAAGTAGTTTCAAGAACTGAAGTTGATGGTGGAAAGT
GGATACAAACCCAAAGAGTTTGATGTTGATCAAGATTGTCACCTGGTTGCTTCAGGGCTGG
AAAGGTAGAATTGTATACGGTTTCATCTATTTGGGTTCTTTCTTTTCTATGTGGGCAGA
GCAACTAGGGTTTTGATCATGGATCCAAACTTCTCTGAATCTCTAAACGGCTTTGAGTAT
TTTGATGGTAACCTAATTTGCTTACTGATCCAATGGAAGATCAGTATCCACCACCATCT
GATACTCTGTTGAAATACGTGAGTGAGATTCTTATGGAAGAGAGTAATGGAGATTATAAG
CAATCTATGTTCTATGATTCATTGGCTTTACGAAAACTGAAGAAATGTTGCAGCAAGTC
ATTACTGATTCTCAAATCAGTCCTTTAGTCCTGCTGATTTCATTGATTACTAATTCTTGG
GATGCAAGCGGAAGCATCGATGAATCGGCTTATTCGGCTGATCCGCAACCTGTGAATGAA
ATTATGGTTAAGAGTATGTTTAGTGATGCAGAATCAGCTTTACAGTTTAAAGAAAGGGGT
GAAGAAGCTAGTAAATCCTTCCCAATAGTGATCAATGGGTATCAATCTGGATATCGAG
AGATCCGAAAGGCGGATTCGGTTAAAGAAGAGATGGGATTGGATCAGTTGAGAGTTAAG
AAGAATCATGAAAGGGATTTTGAGGAAGTTAGGAGTAGTAAGCAATTTGCTAGTAATGTA
GAAGATAGTAAGGTTACAGATATGTTTGATAAGGTTTTCCTTCTTGACGGTGAATGCGAT
CCGCAACATTGTTAGACAGCGAGATTCAAGCGATTCCGAGTAGTAAGAACATAGGAGAG
AAAGGGAAGAAGAAGAAGAAGAAGAGTCAAGTGGTTGATTTTCGTACACTTCTCACT
CATTGTGCACAAGCCATTTCCACAGGAGATAAAACCACGGCTCTTGAGTTTCTGTTACAG
ATAAGGCAACAGTCTTCGCCTCTCGGTGACGCGGGGCAAAGACTAGCTCATTGTTTCGCT
AACGCGCTTGAAGCTCGTCTACAGGGAAGTACCGGTCCTATGATCCAGACTTATTACAAT
GCTTTAACTTCGTCGTTGAAGGATACTGCTGCGGATACAATTAGAGCGTATCGAGTTTAT
CTTTCTTCGCTCCGCTTTGTTACCTTGATGTATTTCTTCTCCATCTGGATGATTCTTGAT
GTGGCTAAAGATGCTCCTGTTCTTCATATAGTTGATTTTGGGATTCTATACGGGTTTCAA
TGGCCGATGTTTATTAGTCTATATCAGATCGAAAAGATGTACCGCGGAAGCTGCGGATT
ACTGGTATCGAGCTTCTCAGTGCGGGTTTCGGCCCCGCGGAGCGAATAGAGGAGACAGGA
CGGAGATTGGCTGAGTATTGTAAACGGTTTAAATGTTCCGTTTGAGTACAAAGCCATTGCG
TCTCAGAACTGGGAAACAATCCGGATAGAAGATCTCGATATACGACCAAACGAAGTCTTA
GCGGTTAATGCTGGACTTAGACTCAAGAACCTTCAAGATGAAAACAGGAAGCGAAGAGAAT
TGCCCGAGAGATGCTGTCTTGAAGCTAATAAGAAACATGAACCCGGACGTTTTCATCCAC
GCGATTGTCAACGGTTTCACTCAACGACCCCTTCTTTATCTCGCGGTTTAAAGAAGCGGTT
TACCATTACTCCGCTCTCTTCGACATGTTTGTGATTTCGACGTTGCCTCGGGATAACAAAGAG
AGGATTAGGTTTCGAGAGGGAGTTTACGGGAGAGAGGCTATGAACGTGATAGCGTGCGAG
GAAGCTGATCGAGTGAGAGGCCCTGAGACTTACAGGCAATGGCAGGTTAGAATGGTTAGA
GCCGGGTTTAAAGCAGAAAACGATTAAAGCCTGAGCTGGTAGAGTTGTTTAGAGGAAAGCTG
AAGAAATGGCGTTTACCATAAAGACTTTGTGGTTGATGAAAATAGTAAATGGTTGTTACAA
GGCTGGAAGGTCGAACCTCTCTATGCTTCTTCTTGTGGGTTTCTGCTAG
>G987 Amino Acid Sequence (domain in AA coordinates: 428-432, 704-708)
MGSYSAGFPGLDWFDFPGLNGSYLNDQPLLDIGSVPPPLDPYPQQNLASADDFSDSV
LKYISQVLMEEDMEDKPCMFHDALSQAAEKSLYEALGEKYPVDDSDQPLTTTTSLAQLV
SSPGSSYASSTTTSSDSQWSFDCLENNRPSSWLQTPIPSNFIFQSTSTRASSGNAVFG
SSFSGDLVSNMFNDTDLALQFKKGMEASKFLPKSSQLVIDNSVPNRLTGKKSHWREEH
LTEERSKKQSAIYVDETDELTDMDNLIIFGEAKEQPVCILNESFPKEPAKASTFSKSPK
GEKPEASGNSYTKETPDLRMTLVSCAQAVSINDRRTADELLSRIQHSSSYGDGTERLAH
YFANSLEARLAGIGTQVYTALSSKKTSTSDMLKAYQTYISVCPFKKIAIIFANHSIMRLA
SSANAKTIHIIDFGISDGFQWPSLIHRLAWRRGSSCKLRITGIELPQRGFRPAEGVIETG
RRLAKYQCQFNIPFEYNIAIAQKWSIKLEDLKLKEGEFVAVNSLFRFRNLLDETVAHVSP
RDTVLKLRIRKIPDVFIPIGILSGSYNAPFFVTRFREVLPHYSSLFDMCDTNLTREDPMRV
MFEKEFYGREIMNVVACEGTERVERPESYKQWQARAMRAGFRQIPLEKELVQKLKLMVES
GYKPKFEFDVDQDCHWLLQGWKGRIVYGSSIWVPPFFVYVGRATRVLIMDPNFSESLNGFEY
FDGNPNLLTDPMEDQYPPPSDTLLKYVSEILMEESNGDYKQSMFYDSLALRKTEMLQQV
ITDSQNQSFSFSPADSLITNSWDASGSIDESAYSADPQPVNEIMVKSMFSDAESALQFKGV
EEASKFLPNSDQWVINLDIERSERRDSVKEEMGLDQLRVKKNHERDFEEVRSSKQFASNV
EDSKVTDMFDKVLILLDGECDPQTLLEDSEIQAIRSSKNIGEKGKKKKKKKSQVDFRLLT

HCAQAISTGDKTTALEFLLQIRQSSPLGDAGQRLAHCFANALEARLQGSTGPMIQTYYN
 ALTSSLKDTAADTIRAYRVYLSSSPFVTLMYFFSIWMILDVAKDAPVLHIVDFGILYGFQ
 WPMFIQSISDRKDVPRKLRLITGIELPQCGFRPAERIEETGRRLAEYCKRFNVPFEYKAIA
 SQNWETIRIEDLDIRENEVLAVNAGLRLKLNQDETGESENCPRDAVLKLRNMNPDVFIH
 AIVNGSFNAPFFISRFKEAVYHYSALFDMFDSTLPRDNKERIRFEREFYGREAMNVIACE
 EADRVERPETYRQWQVRMVRAGFKQKTIKPELVELFRGKLKKWRYHKDFVVDENSKWLLQ
 GWKGRTLYASSCWVPA*

>G993 (6..1091)

CAAATATGGAATACAGCTGTGTAGACGACAGTAGTACAACGTCAGAATCTCTCTCCATCT
 CTACTACTCCAAAGCCGACAACGACGACGAGAGAAGAACTCTCTCTCCGCCGCGACGT
 CGATGCGTCTCTACAGAATGGGAAGCGCGGAAGCAGCGTCGTTTGGATTACAGAGAACG
 GCGTCGAGACCGAGTCACGTAAGCTTCCTTCGTCGAAATATAAAGCGGTTGTGCCTCAGC
 CTAACGGAAGATGGGGAGCTCAGATTTACGAGAAGCATCAGCGAGTTTGGCTCGGTACTT
 TCAACGAGGAAGAAGAAGCTGCGTCTTCTTACGACATCGCCGTGAGGAGATTCCGCGGCC
 GCGACGCCGTCCTAATCTCAAATCTCAAGTTGATGGAACGACGCCGAATCGGCTTTTC
 TTGACGCTCATTCTAAAGCTGAGATCGTGATATGTTGAGGAAACACACTTACGCCGATG
 AGTATTGAGCAGAGTAGACGGAAGTTTGTAAACGCGACGGAACGCTCTGGGTTGGAGA
 CGGCGACGTACGGAACGACGCTGTTTTGAGAGCGCGTGAGGTTTTGTTTCGAGAAGACTG
 TTACGCCGAGCGACGTGCGGAAGCTGAACCGTTTAGTGATACCGAAACAACGCGGAGA
 AGCATTTCCTGTTACCGCGATGACGACGCGGATGGGGATGAATCCGTCTCCGACGAAAG
 GCGTTTTGATTAACTTGAAGATAGAACAGGGAAGTGTGGCGGTTCCGTTACAGTTACT
 GGAACAGCAGTCAAAGTTACGTGTTGACCAAGGGCTGGAGCCGTTTCGTTAAAGAGAAGA
 ATCTTCGAGCCGGTGATGTGGTTTTGTTTCGAGAGATCAACCGGACCAGACCGGCAATTGT
 ATATCCACTGGAAGTCCGGTCTAGTCCGGTTCAGACTGTGGTTAGGCTATTCCGAGTCA
 ACATTTTCAATGTGAGTAACGAGAAACCAAACGACGTCGAGTAGAGTGTGTTGGCAAGA
 AGAGATCTCGGGAAGATGATTTGTTTTTCGTTAGGGTGTTCCAAGAAGCAGGCGATTATCA
 ACATCTTGTCACAAATCTTTTTTTTTGTTTTTTCTTCAATTGTTTCTCCTTTTTTCA
 ATATTTTGTATTGAAATGACAAGTTGTAAATTAGGACAAGACAAGAAAAAATGACAACTA
 GACAAAATAGTTTTTGTATAAAAAAAAAAAAAAAAAAAAAA

>G993 Amino Acid Sequence (domain in AA coordinates: 69-134)

MEYSCVDDSTTSSELSISTTPKPTTTTEKLLSPPATSMRLYRMGSGSSSVLDSENGV
 ETESRKLPSKSKYKGVVPQPNRWGAQIYEKHQRVWLGTFNEEEEAASSYDIAVRRFRGRD
 AVTNFKSQVDGNDAESAFDLAHSKAEIVMLRKHTYADEFEQSRKRFVNGDGKRSGLTA
 TYGNDAVLRAREVLFEKTVTPSDVGKLNRLVIPQHAKEKHFPLPAMTTAMGMNPSPTKGV
 LINLEDRGKGVWRFRYSYWNSSQSYVLTGWSRFVKEKNLRAGDVVCFERSTGPDRQLYI
 HWKVRSSPVQTVVRLFGVNI FNVSNKPNDAVECVGKKRSREDDLFSLGCSKKQAIINI
 L*

>G681 (1..804)

ATGGGGAGGACGACATGGTTCGACGTCGACGGGATGAAGAAAGGAGAGTGGACGGCAGAG
 GAAGACCAGAAGCTCGGCGCTTACATCAACGAGCATGGCGTTTGTGATTGGCGTTCCCTC
 CCCAAAGAGCTGGTTTGAGAGATGTGGAAAGAGCTGCAGATTAAGTGCGCTTAATAT
 CTAAAGCCTGGGATTAGAAGAGGCAAATTCCTCCTCAAGAAGAAGAAGAAATCATCCAA
 CTTTCATGCTGTTCTCGGAAACAGGTGGGCGAGCCATGGCGAAGAAGATGCAGAATCGAACA
 GACAATGATATCAAGAACCATTGGAACCTTGTCTCAAGAAAAGACTTTCGAGAAAGGGA
 ATCGACCCTATGACCCACGAGCCCATCATCAACACCTCACCGTCAATACCACTAACGCA
 GATTGTGGTAACTCTTCCACCACGACGTCCCGTCGACGACGGAAGCTCTCCTTCCTCC
 GGCTCGTCTCGTCTTCTTAACAAACTCGCCGCAGGTATCTCATCTAGACAACATAGTCTC
 GATAGGATCAAGTAACATCTTGTCTGAATCAATAATCGAAAGCAGTGATCAAGCAAAAGAG
 GAAGAAGAAAAAGAAGAAGAAGAAGAAGAAGAGATTCAATGATGGGTGAGAAGATTGAC
 GGTAGTGAAGGAGAAGATATTCAGATTTGGGGCGAGGAGGAAGTTAGGCGTTTAAATGGAG
 ATTGATGCAATGGATGATGACGAGATGACTTCGTACGACGCTGTCATGTACGAGAGTAGT
 CACATACCTGATCATCTCTTTTGACTTAATATAGTGTGACTGTGTGAGTGCATGCATGTT
 >G681 Amino Acid Sequence (domain in AA coordinates: 14-120)

MGRTTWFDVDMKKGEWTAEEDQKLGAYINEHGVCWRSRSLPKRAGLQRCGKSCRLRWLNY
 LKPGIRRGKFTPEEEEEIIQLHAVLGNRWAAAMAKKMQNRDNDIKHNWNSCLKKRLSRKG
 IDPMTHEPIIKHLTVNTTNADCGNSSSTTSPSTTESPSSSGSSRLLNKLAAAGISSRQHS
 DRIKYILSNSIISSDQAKEEEEEEEEEERDSMMGQKIDGSEGEDIIQIWGEEVEVRLME

IDAMDMYEMTSYDAVMYESSHILDHLF*

>G1482 (1..996)

ATGAAGATCAGGTGCGACGCTGCGATAAAGAAGAAGCGTCGGTGTTTTGCACGGCCGAC
GAAGCATCTCTGCGGCGCTGCGACCACCAAGTCCACCACGCTAACAAACTCGCCTCT
AAACATCTCCGTTTTCTCTCCTTTATCCTTCTTCTTCCAACACCTCCTCTCTCTGCG
GACATCTGTGAGGATAAAAAAGCTCTGTTGTTCTGTCAACAAGATAGAGCTATTTTATGC
AAAGATTGCGATTATCGATCCACGCTGCGAACGAACACACAAAGAAACACGATAGGTTT
CTTCTTACAGGGGTTAAGCTCTCTGCAACATCGTCTGTTTACAAACCTACTTCGAAATCT
TCTTCTTCTTCTTCAAGCAACCAAGATTTCTCTGTCCTTGATCATCAATCTCTAATCCT
CCTCCTCTCAAGAAACCTCTCTCAGCTCCTCCTCAGAGCAACAAGATCCAACCTTTTCG
AAGATCAACGGCGGTGATGCGTCGGTGAATCAGTGGGGATCCACAAGCAGATTCTGAG
TATTGATGGATACGTTACCTGGTTGGCAGCTTGAGGATTTCTCGATTCTCTCTTCTCT
ACTTATGGTTTCTCTAAGAGTGGTGATGATGATGGAGTGTTACCATATATGGAACCAGAA
GATGACAACAACACTAAGAGAAAACAACAACAACAACAACAACAACAACAATACAGTG
TCACTTCCATCTAAGAATTTAGGGATTTGGGTCCCTCAGATTCCACAACTCTTCTCTCT
TCATACCCAAATCAATACTTTTCTCAAGACAACAACATACAGTTTGGGATGTACAACAAA
GAAACATCACCAGAAGTAGTGCTTTTGTCTCCAATACAAAACATGAAACAACAAGGACAG
AACAAACAAGAGATGGTATGATGATGGTGGCTTCACTGTCCACAGATCACTCCTCCTCT
CTTCTCTCTAATAAAAAGTTTAGATCTTTCTGGTAA

>G1482 Amino Acid Sequence (domain in aa coordinates: 5-63)

MKIRCDVCDKEEASVFCTADEASLCGGCDHQVHANKLASKHLRFSLYPSSNTSSPLC
DICQDKKALLFCQQDRAILCKDCDSSIHAANEHTKKHDFLLTGVKLSATSSVYKPTSKS
SSSSSSNQDFSVPGSSISNPPPLKKPLSAPPQSNKIQPFKINGGDASVNQWGSTSTISE
YLMDTLPGWHVEDFLDSSLPTYGFSKSGDDDGVLPMPEPDNNTKRNNNNNNNNNNTV
SLPSKNLGIWVPQIPQTLPSYPNQYFSQDNNIQFGMYNKETSPEVVSFAPIQNMKQGGQ
NNKRWYDDGGFTVPQITPPPLSSNKKFRSFW*

>G225 (157..441)

CTCTCTCTCTCACTCTTTTCTTTTCCGAGAACCCAACAAAAAAGCTACTATTAATCC
TTCCCTCGTGAGGAAATCATTTCTTCTGTTTCTCGAGATTTATTCTCTTTCTCTCTCT
CTTTCTCTGTGTGTTTCGTGTCTTCAGATTAGTTTCGATGTTTCGTTTCAGACAAGCGGAA
AAAATGGATAAACGACGACGAGAGACAGAGCAAAGCCAAGGCTTCTTGTTCCGAAGAGGTG
AGTAGTATCGAATGGGAAGCTGTGAAGATGTGAGAAGAAGAAGATCTCATTCTCTCGG
ATGTATAAACTCGTTGGCGACAGGTGGGAGTTGATCGCCGGAAGGATCCCGGACGGACG
CCGAGGAGATAGAGAGATATTGGCTTATGAAACACGGCGTCGTTTTTGCCAACAGACGA
AGAGACTTTTTTAGGAAATGATTTTTTTTGTGTTGGATTAAAAGAAAATTTCTCTCTCT
AATTACAAAGACAAGAAAAAAGGAAATGTACCTGTCTTGAATTACTATTTTGAATGT
ATAATTATCTATATATATAAGAAGAAAAAATTGCTTAGGAATTT

>G225 Amino Acid Sequence (domain in AA coordinates: 39-76)

MFRSDKAEMDKRRRRQSKAKASCSEEVSSIEWEAVKMSEEEEDLISRMYKLVGDRWELI
AGRIPGRTPEEIERYWLMKHGVVFANRRRDFFRK*

>G226 (10..348)

CCAGTAGTTATGGATAATACCAACCGTCTTCGTCTTCGTGCGGTCCCAGTCTTAGGCAA
ACTAAGTTCACTCGATCCCGATATGACTCTGAAGAAGTGAGTAGCATCGAATGGGAGTTT
ATCAGTATGACCGAACAAGAAGAAGATCTCATCTCTCGAATGTACAGACTTGTCGGTAAT
AGGTGGGATTTAATAGCAGGAAGAGTCGTAGGAAGAAAGGCAATGAGATTGAGAGATAC
TGGATTATGAGAACTCTGACTATTTTCTCACAACGACGACGTCTTAATAATTCTCCC
TTTTTTTCTACTTCTCTCTTAATCTCCAAGAAAATCTAAAATTGTAAAGAAATCAAAT
AAAAGCTTTCAATCAATAAAGTAGAACAAATCTTGAATGTCTTCTCA

>G226 Amino Acid Sequence (domain in AA coordinates: 28-78)

MDNTNRLRLRRGPSLRQTKFTRSRDYSEEVSIEWEFISMTEQEEDLISRMYRLVGNRWD
LIAGRVVGRKANIEIRYWIMRNSDYFSHKRRRLNNSPFFSTSPNLQENLKL*

>G9 (81..1139)

GTGTTTCTTCTTTCTGCTAAAGGTTATAATTTTTGTTTCTTGGTTTGGTGAGAATCTTC
AAGAACTGAAACAAAGAAAATGGATTCTAGTTGCATAGACGAGATAAGTTCTCCACTT
CAGAATCTTTCTCCGCCACCACCGCCAAGAAGCTCTCTCCTCCTCCCGCGGCGGCTTAC
GCCTCTACCGGATGGGAAGCGGCGGGAGCAGCGTCGTGTTGGATCCCGAGAACGGCCTAG
AGACGGAGTACGAAAGCTACCATCTTCAAAATACAAAGGTGTTGTTCTCAGCCTAACG

GAAGATGGGGAGCTCAGATCTACGAGAAGCACCAACGAGTATGGCTCGGGACTTTCAACG
 AGCAAGAAGAAGCTGCTCGTTCCTACGACATCGCAGCTTGTAGATTCCGTGGCCGCGACG
 CCGTCGTCAACTTCAAGAACGTTCTGGAAGACGGCGATTTAGCTTTTCTTGAAGCTCACT
 CAAAGGCCGAGATCGTCGACATGTTGAGAAAAACACTTACGCCGACGAGCTTGAACAGA
 ACAATAAACGGCAGTTGTTTCTCTCCGTCGACGCTAACGGAAAACGTAACGGATCGAGTA
 CTACTCAAAACGACAAAAGTTTTAAAGACGTGTGAAGTCTTTTCGAGAAGGCTGTTACAC
 CTAGCGACGTTGGGAAGCTAAACCGTCTCGTGATACCTAAACAACACGCCGAGAAACACT
 TTCCGTTACCGTCACCGTCACCGGCAGTGAATAAGGAGTTTGTATCAACTTCGAAGACG
 TTAACGGTAAAGTGTGGAGGTTCCGTTACTCATACTGGAACAGTAGTCAAAGTTACGTGT
 TGACCAAGGGATGGAGTCGATTTCGTCAAGGAGAAGAATCTTCGAGCCGGTGATGTTGTTA
 CTTTCGAGAGATCGACCGGACTAGAGCGGCAGTTATATATTGATTGGAAAGTTCGGTCTG
 GTCCGAGAGAAAACCCGGTTCAGGTGGTGGTTCGGCTTTTCGAGTGTATATCTTTAATG
 TGACCACCGTGAAGCCAAACGACGTCGTGGCCGTTTTCGGTGGAAAGAGATCTCGAGATG
 TTGATGATATGTTTTCGCTTACGGTGTTCGAAGAAGCAGGCGATAATCAATGCTTTGTGAC
 ATATTTCTTTTCCGATTTTATGCTTTTCGTTTTTAATTTTTTTTTTTTGTCAAGTTGTGT
 AGGTTGTGATTATGCTAGGTTGTATTTAGGAAAAGAGATAAGACC

>G9 Amino Acid Sequence (domain in AA coordinates: 62-127)
 MDSSCIDEISSSTSESFSAATAKKLSPPPAALRLYRMGSGGSSVLDPENGLTESRKL
 PSSKYKGVVPQPNRWGAQIYEKHQVWLGTFFNEQEEAARSYDIAACRFRGRDAVNVFN
 VLEDGDLAFLEAHSAEIVDLRKHTYADELEQNNKRLFLSVDANGKRNGSSTTQNDKV
 LKTCEVLFEKAVTPSDVGKLNRLVLPKHAEKHFPLPSPSPAVTKGVLINFEDVNGKVWR
 FRYSYWNSSQSYVLTKGWSRFVKEKNLRAGDVVTFERSTGLERQLYIDWKVRS GPENPV
 QVVVRLFGVDIFNVTTVKPNDVVAVCGGKRSRDVDDMFALRCSKKQAIINAL*

>G1040 (51..863)
 CTTTGATCTCCACTATTTAAGTAGACAAGAATCATAAAGAAAATAGTGAGATGATGATGT
 TAGAGTCAAGAAACAGTATGAGAGCTTCAAACTCAGTCCCAGATCTGTCTCTTCAGATCA
 GTCTTCTTAACATATCACGCCGAAAACCTCTTCACGGCGGTGACCGAGCTCCACAAGCA
 GTGATTCTGGAAGCAGCCTCAGTGACCTGAGCCATGAGAACAACCTCTTCAACAAACCTC
 TCTTGAGCTTAGGATTTGACCATCATCATCAAAGGCGCTCAAACATGTTCCAACCTCAAA
 TCTACGGTCGAGATTTCAAGAGAAGCTCATCATCAATGGTTGGTCTTAAACGAAGCATTCT
 GTGCTCCAAGAATGAGATGGACTTCTACTCTTCATGCTCACTTCGTCATGCTGTTCAAC
 TTCTTGGCGGCCATGAAAGAGCAACGCCATAATCAGTGTGGAGCTCATGAATGTGAAGG
 ATCTAACCTTAGCTCATGTCAAGAGTCACTTGACAGATGTATAGAACAGTGAAATGCACTG
 ATAAAGGATCACCAGGAGAAGGAAAGGTAGAGAAAGGAGCAGAGCAGAGGATAGAGGACA
 ATAATAATAATGAAGAGCTGATGAAGGAAGTACACAAATTCGCCAAACTCATCATCTG
 TGCAAAAGACCCAAAGAGCTTCATGGTCATCGACAAAGGAAGTATCTAGGAGCATATCTA
 CACAAGCATATTTCTCACTTGGGAACAACATCATCACTAAGGCCAATGAAGAGAAAGAGG
 ATACCAACATTCATCTCAATTTGGATTTACATTGGGCGGCCTAGTTGGGGGATGGAATA
 TGCGGAACCTCCAGTGATTTAACCTTCTCAAGTGCTAATTGCCTTAAGCTACAACAAA
 TAAGTCAGCTTAGGTTACAGTTTTAAACATAATTTTAACTTGTTTTGATCATATGAGCTT
 CGGAAGAATCATATTATCATCATATATGAACCTCTTCCAAGAATGTCTATGAGTTTTT
 TGATATGTATAATCAAGAGAATCGTTTGAAGTAAAAA

>G1040 Amino Acid Sequence (domain in AA coordinates: 109-158)
 MMLLESNRSMRASNSVPDLQLISLPNYHAGKPLHGGDRSSTSSDSGSSLSDLSHENNPF
 NKPLLSLGLFDHHRQRNSNMFPQIYGRDFKRSSSSMVLKRSIRAPRMRWTSTLHAHFVH
 AVQLLGGERATPKSVLELMNVKDLTLAHVKSHLQMYRTVKCTDKGSPGEGKVEKEAEQR
 IEDNNNNNEADEGTDINSFNSSSVQKTQRASWSSTKEVSRISISTQAYSHLGTTHHTKANE
 EKEDTNIHLNLDFTLGGVLGGWNMRNPPI*

>G2114 (64..1311)
 ATAAAACGAAACCTATACATATAAACTAAGAGCGAGAAAGACAGCTAGAGAGAGAGAGA
 GAGATGAAGAAATGGTTGGGATTTTCAATTGACACCTCCTTTGAGAATCTGCAATAGTGAA
 GAAGAAGAACTTAGGCATGACGGTTCGATGTTTGGAGATATGATATTAACCTTTGATCAT
 CATCATCATGATGAAGACGTTCCAAAGGTGGAAGATCTCCTCTCAAACCTCTCATCAAACC
 GAGTATCTATAAACCATAAACCAACCAATGTCAACTGCACCACTGTGGTTAACAGGTTA
 AACCACCCCGTTACCTTCTCCACGACCAACCGTAGTTACACCACATTACCCGAACCTA
 GATCCGAACCTTAGCAATGATTATGGAGGTTTGGAGGGTTCGGTTCGGTCTCGGTTTTTC
 AAATCTTGGTTAGAGCAAGGCACTCCAGCATTCCCACTCTCGAGTCAATACGTTACTGAA

GAGGCTGGTACGAGCAATAATATTAGTCATTTTAGTAACGAAGAGACTGGTTATAACACC
 AATGGCTCAATGCTATCATTGGCTTTGAGCCATGGGGCTTGTCTGATTTGATCAACGAA
 TCGAATGTATCCGCACGGGTCGAAGAACCGGTTAAGGTAGATGAGAAGCGGAAGAGATTG
 GTTGTTAAACCTCAGGTAAAGGAATCCGTTCTCGGAAGTCGGTTGATAGTTATGGACAA
 AGAAGTTCTCAGTATCGTGGAGTTACAAGGCATAGATGGACAGGGAGATATGAAGCTCAC
 TTATGGGATAATAGCTGTAAGAAGGAGGGACAGACAAGGAGAGGAAGACAAGTGTATCTT
 GGAGGGTATGATGAGGAGGAGAAAGCAGCGAGGGCATATGATTTAGCGGCTCTGAAGTAT
 TGGGGTCTTACCCTCACTTAAATTTCCCTTTGAGTAATTACGAAAAGGAGATCGAGGAA
 CTCAATAACATGAATCGGCAAGAATTTGTTGCCATGTTGAGGAGGAATAGCAGCGGGTTT
 TCGAGGGGAGCTTCCGTGTATAGAGGAGTTACAAGGCATCATCAACATGGAAGGTGGCAA
 GCCAGAATTGGAAGAGTTGCTGGAAACAAGGACTTGTACCTTGAACATTTAGCAGCGAA
 GAAGAAGCAGCGGAGGCGTACGATATCGCGGCAATTAAATTCAGAGGCCTAAACCTGTGA
 ACCAATTTGATATAAATAGATATGACGTGAAGAGGATATGTTCAAGCTCAACGATTGTT
 GATAGCGACACCGCCAAACATTCTCCACAGCTCTGGCGCCGCCACTAACCACACCG
 TAAACTCCTCGCCGGAGAGACTATTCCACGTACGGTTGGTTGAGGAAATAAGTTCGTC
 CAGTCTGTTTAAATCATTTATGGTTTAAATAACATATATTCCTAAGTAATTGAGGCCGGTC
 TACATATATACAACTTTTTTAGCAAATTAAGTTATCAGAATCCACTATATATTATTCTCT
 >G2114 Amino Acid Sequence (conserved domain in AA coordinates:221-297, 323-393)

MKKWLGLFSLTPPLRICNSEEEELRHDGSDVWRYDINPDHHHDEEDVPKVEDLLNSHQTE
 YPINHNQTNVNCCTTVNRLNPPGYLLHDQTVVTPHYPNLDPNLSNDYGGFERVGSVSVFK
 SWLEQGTAPFLSSHYVTEEAGTSNNISHFSNEETGYNTNGSMLSLALSHGACSDLINES
 NVSARVEEPVKVDEKRLVVKPQVKESVPRKSVDSYGQRTSQYRGVTRHRWTRGRYEHL
 WDNSCKKEGQTRRGRQVYLGVDDEEKAARAYDLAALKYWGPTTHLNFPLSNYEKEIEEL
 NNMNRQEFVAMLRNSSGFSRGASVYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFTSTQE
 EAAEAYDIAAIKFRGLNAVTFNFDINRYDVKRICSSSTIVDSQAKHSPTSSGAGH*
 >G450 (65..751)

GAGTTATCGAGAGAGAGAGAAAAACATATTTCTGATTTAAGACATATATAGACAGCAAGAAG
 AGATATGAACCTTAAGGAGACGGAGCTTTGTCTTGGCCTCCCCGGAGGCACTGAAACCGT
 TGAAAGTCCGGCCAAGTCGGGTGTTGGGAACAAGAGAGGCTTCTCCGAGACCGTTGATCT
 CAAACTTAATCTTCAATCTAACAAACAAGGACATGTGGATCTCAACACTAATGGAGCTCC
 CAAGGAGAAGACCTTCTTAAAGACCTTCTAAGCCTCCTGCTAAAGCACAAAGTGGTGGG
 TTGGCCACCGGTGAGGAACCTACCGGAAAAATGTTATGGCTAATCAGAAGAGCGGCGAAGC
 AGAGGAGGCAATGAGTAGTGGTGGAGGAACCGTCGCCTTTGTGAAGGTTTCCATGGATGG
 AGCTCCTTATCTTCGGAAGGTTGACCTCAAGATGTACACCAGCTACAAGGATCTCTCTGA
 TGCCTTGGCCAAAATGTTTCAGCTCCTTTACCATGGGGAGTTATGGAGCACAAGGGATGAT
 AGATTTTCATGAACGAGAGTAAAGTGATGGATCTGTTGAACAGTTCTGAGTATGTTCCAAG
 CTACGAGGACAAAGATGGTGAAGTGGATGCTCGTTGGTGATGTCCCTTGGCCGATGTTTGT
 CGAGTCATGCAAAACGTTTGCGCATAATGAAAGGATCCGAAGCAATTGGACTTGCTCCAAG
 AGCAATGGAGAAGTTCAAGAACAGATCATGAACAAAAAAGAGGACAAATATGCATTG
 ATTTTTTTTTTTTTTGGTATTGTTATGATCATGTGTTTAAATTTAAATATAGGAAGGATA
 TAGGAAAAATATAATTGTTTACAAAAAATAACTTTAAATATGTCTTTTTTTTTTTTTTTGA
 AATTAGTCTGTGTTTTTGTTCATCTCTTAATTAGTAGAAATCATTTTTTAATATGTAA
 TTGTGATAGTAAATCTATAGAGTTTCGTA

>G450 Amino Acid Sequence (domain in AA coordinates: TBD)
 MNLKETELCLGLPGGTETVESPAKSGVGNKRGFSETVDLKLNLQSNKQGHVDLNTNGAPK
 EKTFLKDPSPKPPAKAQVVGWPPVRNRYRKNVMANQKSGEAEEMSSGGGTVAFVKVSMGDA
 PYLRKVDLKMYSYKDLSDALAKMFSSFTMGSYGAQGMIDFMNESKVMDDLNSSEYVPSY
 EDKGDWMLVGDVPWPMFVESCKRLRIMKGSEAIGLAPRAMEKFNRS*

>G584 (40..1809)
 AAAAAGTCTTCTCTTTTATAACTACGTCAGAGAACTGTTATGTCTCCGACGAATGTTCAA
 GTAACCGATTACCATCTCAACCAATCAAAAACGGATACAACAAATCTCTGGTCAACCGAC
 GACGATGCATCGGTAATGGAAGCTTTCATCGCGCGCGGCTCCGATCATTCTTCTTTTTT
 CCTCCACTTCTCTCTCTCTCTCTCTCAAGTCAACGAAGATAATCTCCAGCAACGTCTC
 CAAGCTTTAATCGAAGGAGCAACGAGAACTGGACTTACGCCGTGTTCTGGCAATCATCT
 CACGGTTTCGCGGAGAGAACAACAACAACAACACAGTGTGTTAGGTTGGGGAGAT
 GGTTATTACAAGGAGAAGAAGAGAAGTCTAGAAAAGAAGAAATCAAATCCAGCTAGTGCA
 GCTGAACAAGAGCATCGTAAGAGAGTGATTAGAGAGCTCAACTCTTAATCTCCGGTGGT

GTAGGAGGAGGAGATGAAGCTGGAGATGAAGAAGTTACAGATACTGAATGGTTCTTCTTA
 GTTTCATGACACAGAGCTTTGTCAAGGGTACTGGTTTACCTGGTCAAGCTTTCTCAAAT
 TCAGACACGATTTGGTTATCTGGTTCTAATGCTTTAGCTGGATCAAGTTGTGAGAGAGCT
 CGTCAAGGTGAGATTTATGGGTTACAAACAATGGTGTGTGTAGCGACAGAGAATGGTGTG
 GTTGAGCTTGGTTCGTCGGAGATTATTCATCAAAGTTCAGATCTTGTGTATAAAGTTGAC
 ACCTTTTTCAATTTTAACAATGGTGGTGGTGAATTTGGTTCTTGGGCGTTTAAATTTGAAT
 CCAGATCAAGGAGAGAATGATCCAGGTTTGTGGATTAGTGAACCTAATGGTGTGACTCT
 GGTCTTGTAGCTGCTCCGGTGATGAATAATGGTGGAAATGACTCAACTTCTAATTTCTGAT
 TCTCAACCAATTTCTAAGCTTTGTAATGGAAGCTCTGTTGAAAACCTTAACCTTAAAGTT
 CTGAAATCTTGTGAAATGGTGAATTTCAAGAATGGGATTGAGAATGGTCAAGAAGAAGAT
 AGTAGTAATAAGAAGAGATCACCGGTTTCAATAATGAAGAAGGGATGCTTCTTTTACC
 TCTGTTCTTCCATGTGACTCGAATCACTCTGATCTTGAAGCTTCAGTGGCTAAAGAAGCT
 GAGAGTAACAGAGTTGTGGTTGAACCGGAGAGAAGAAACCGAGGAAACGAGGGAGAAAACCG
 GCGAATGGAAGAGAAGAGCCTTTGAATCATGTAGAGGCAGAGAGACAGAGAAGAGAGAAG
 TTGAATCAGAGATTCTATTCTTTAAGAGCTGTGGTTTCTAATGTGTCTAAGATGGATAAA
 GCTTCTCTATTAGGAGATGCTATTTCTGATATCAGTGAGCTTAAGTCTAAGTTGCAAAAG
 GCTGAATCTGATAAAGAAGAGTTGCAGAAGCAGATTGATGTGATGAATAAAGAAGCGGGA
 AATGCCGAAAAGTTCCGGTAAAGATCGAAAATGTTTGAATCAAGAATCGAGTGTGTTGATA
 GAGATGGAGGTTGATGTGAAGATTATTGGTTGGGATGCAATGATAAGGATTCAATGTAGT
 AAGAGGAATCATCCTGGTGTCTAAGTTCATGGAAGCACTTAAGGAGTTGGATTGGAAGTG
 AATCATGCGAGTTTATCGGTAGTGAATGATCTTATGATCCAACAAGCGACTGTGAAAATG
 GGGAATCAGTTTTTTCACGCAAGATCAACTCAAGGTTGCTCTAACGGAGAAAAGTTGGAGAA
 TGTCCATGAATTGAAGTCAGCATCTTTAGGGCTAATACACCGGAGAATACTGCGAAAAGT
 CGAAAACAACGATCATAGTATAAGCCGCGGTAAAAAGTGTAAACCTTTTACACAAGTTT
 CTCTAGTGAATGTATGTAACCTCTATTGTGTGAAGGTAATTTTGTAGTACCCACTTGT
 TGCTATTGAATGCTTGTGTAGAGGATTCTTAGTGTAGTATATGATTAGGTTGGGGTTTG
 TTGTTTCATGAGATAAATAAATGTGTTTGATCAATGGTTAAGTCTTTGGTTTGTGGTGT
 ATGTATGTAAATAAGGCTTTTGTAGAAATAAGACAAATGGGACTGAAGTTGGAGTTTAA
 AA

>G584 Amino Acid Sequence (domain in AA coordinates: 401-494)
 MSPTNVQVTDYHLNQSKTDTTNLWSTDDDDASVMEAFIGGSDHSSLFPPLPPPLPQVNE
 DNLQRLQALIEGANENWTVYVFWQSSHGAFAGEDNNNNNTVLLGWGDGYKGEEEKSRKK
 KSNPASAAEQEHRKRVIRELNSLISGGVGGGDEAGDEEVDTEWFFLVSMQSFVKGTGL
 PGQAFNSNDTIWLSGSNALAGSSCERARQGQIYGLQTMVCVATENGVVVELGSSELIHQSS
 DLVDKVDTFNFFNNGGGEFGSWAFNLNPDQGENDPGLWISSEPNVDSGLVAAPVMNNGN
 DSTSNSDSQPIKSLCNGSSVENPNPKVLKSCMVNFKNIGENGQEESSNKKRSPVSNNE
 EGMLSFTSVLPDCSNHSDLEASVAKEAESNRVVVEPEKKPRKRGRKPANGREEPLNHVEA
 ERQRREKLNQRFYSLRAVVPNVSKMDKASLLGDAISYISELKSQKQAEKQKQID
 VMNKEAGNAKSSVKDRKCLNQESSVLIEMEVDVKIIGWDAMIRIQSKRNHPGAKFMEAL
 KELDLEVNHASLSVNDLMIQQATVKMGNQFFTQDQLKVALTEKVGEC*

>G668 (1..1056)

ATGGGAAGACCACCTTGCTGTGAAAAGATTGGAGTGAAGAAAGGGCCATGGACACCAGAG
 GAAGACATCATCTTGGTTTCTTACATCCAAGAACATGGTCTGGAACCTGGAGATCTGTC
 CCAACACACACAGGTTTAAAGATGTAGCAAGAGCTGCAGATTGAGATGGACTAATTATCTT
 CGACCCGGTATTAAGCGTGGAAATTTTACTGAGCATGAAGAGAAGACAATTGTTTCATCTT
 CAAGCCCTTTTAGGCAACAGATGGGCAGCCATAGCATCATACCTTCCAGAAAGGACAGAC
 AATGATATAAAGAACTATTGGAACACTCACTTGAAGAAGAAGCTCAAAAAGATTAATGAA
 TCTGGTGAAGAAGATAATGATGGTGTCTCTTCATCAAACACTAGTTTCAAAAAGAACCAT
 CAAAGCACTAACAAAGGTCAATGGGAAAGAAGACTTCAGACAGACATTAACATGGCAAAA
 CAAGCTCTTGTGAGGCCTTGTCTTTAGACAAACCATCATCCACTCTTTCATCATCTTCA
 TCATTACCGACACCCAGTAATCACACAACAAACATCCGTAACCTTCTCATCAGCTTTGCTT
 GACCGTTGTATGATCCATCCTCTTCTTCTTCTTACCTACCAACCACTTCAAGCAAC
 ACTACTAATCCATACCCATCAGGGGTATATGCGTCAAGTGCTGAGAACATCGCCCGGTTG
 CTTCAAGATTTTCATGAAAGACACACCCAAGGCTTTAACTTTATCATCTTTCATCTCCGTT
 TCAGAGACTGGACCACTCACTGCTGAGTCTCGGAAGAAGGTGGAGAAGGGTTTGAACAA
 TCTTTCTTCAGCTTCAATTCATGGACGAAACTCAAACTTGACTCAGGAGACAAGCTTC
 TTCCATGATCAAGTGATCAAACCGGAAATAACAATGGACCAAGATCATGGTCTAATATCA

CAAGGGTCTCTGTCTTTGTTTGAGAAATGGTTATTTGATGAGCAAAGCCACGAGATGGTT
GGTATGGCACTAGCAGGACAAGAAGGGATGTTCTAG

>G668 Amino Acid Sequence (domain in AA coordinates: 13-113)
MGRPPCEKIGVVKGPWTPEEDIILVSYIQEHGPGNWRVPTHTGLRCSKSCRLRWNYL
RPGIKRGNFTEHEEKTIVHLQALLGNRWAAIASYLPERTDNDIKNYWNTHLKKLKKINE
SGEEDNDGVSSSNTSSQKNHQSTNKGQWERRLQTDINMAKQALCEALSLDKPSSLSSSS
SLPTPVITQQNIRNFSSALLDRCYDPSSSSSSTTTTTNTNTPYPSGVYASSAENIARL
LQDFMKDTPKALTLSSSSPVSETGPLTAAVSEEGGEGFEQSFFSFNSMDETQNLQTQTSF
FHDQVIKPEITMDQDHGLISQGSLSLFEKWLFEQSHMVMALAGQEGMF*

>G1050 (23..1582)
TTCCCCATTTTCAGAAAATCAAAATGGGTGGTGGTGGTGATACAACAGATACCAATATGAT
GCAGAGAGTTAATCTTCTTCTGGTACATCGTCTTCTTCGATCCCTAAACACAATCTTCA
CTTGAATCCTGCTCTTATCCGCTCTCACCATCACTCCGTCACCCTTTCACCGGAGCTCC
TCCACCGCCGATTCCACCCATTTCTCCTTACTCTCAGATCCCGGCGACTTTACAACCTAG
ACATTCTCGCTCTATGTCGAACCGTCTTCTTCTTCTCCTTTGATTTCATTGCCGCCGT
AAATCCTTCTGCTCCGTCGGTTTCGGTGTTCGGTGGAGGAGAAAACCGGTGCCGGATTAG
TCCTTCGGTTGCCCTCCGTCACCGTTTACGATGTGTCTTCTTAGCTCTAGGAACGCCGG
AGATGGAGAGAATCTACCTCCGAGAAAGTCGCATAGGCGTTCGAATAGTGATGTTACTTT
TGGGTTTAGTTCAATGATGTCTCAGAATCAAAAGTCTCCTCCTTTGAGTTCTTTGGAGAG
ATCGATCTCTGGTGAAGATACATCAGATTGGTCTAATTTGGTGAAGAAAGAACCGAGAGA
AGGCTTCTACAAGGGAAGAAAACAGAGGTTGAAGCAGCTATGGACGATGTTTTCACGGC
TTATATGAATCTTGATAACATTGATGTCTTGAATCTTTTGGAGGTGAAGATGGCAAGAA
TGGGAATGAGAATGTGGAGGAGATGGAGAGTAGTAGAGGTAGTGGTACAAAGAAGACGAA
TGGTGAAGTAGTAGTGATTCTGAAGGAGATAGCAGTGCGAGTGGGAATGTGAAGGTTGC
GTTGAGTTCTTCTTCTCAGGCGTGAAGAGAAGAGCAGGTGGAGATATTGCTCCTACTGG
TAGACATTACAGGAGTGTCTTCTATGGACAGTTGTTTCATGGGGAAGTTGAATTTCCGCCA
CGAATCATCGCTAAAGCTTCCGCCCTTCTTCATCAGCTAAAGTTTCCCCAACCAATTCAGG
TGAAGGGAATTCAAGTGCTTATAGTGTGAATTTGGAACAGTGAGTTTACTGCAGCTGA
AATGAAGAAGATTGCAGCTGATGAGAACTCGCTGAGATTGTAATGGCTGACCCTAAGCG
TGTTAAAAGAATCTTGGCGAACC CGGTATCTGCTGCACGTTCAAAGGAGCGGAAGACGCG
ATACATGGCAGAGTTGGAACACAAGGTGCAGACACTTCAGACTGAAGCTACTACATTATC
GGCTCAGCTCACACATTTGCAGAGAGATTCTATGGGGTTGACAAACCAGAACAGTGAGCT
GAAGTTTCGTCTTCAAGCTATGGAGCAGCAAGCACAACCTCCGCGATGCTCTGTGAGAGAA
ACTGAATGAAGAAGTCCAGCGTTGAAACTGGTGTAGGGGAGCCGAACCGCAGGCAAAG
TGGGAGCAGCAGCAGCGAATCAAAGATGTCACTAAACCCGAGATGTTTCAGCAGCTTAG
CATAAGTCAGTTACAACACCAACAGATGCAGCATTCCAATCAGTGTAGCACAATGAAAGC
AAAGCACACTTCAAACGACTAGGGTAAGTAAACTGCGATCCGCAGTTGTCTAGTTACAT
ATATGATAAGAATCTTTTGTGAGAGTTCTGTTTTTGAAGTTTTTAAAGAAACATATATA
AAGATTATGTCCGGGAAATTTGATCATATTTCTGAAACATACACATATATATAGTGG
TAATGGAGGACTTTCTTTCTGGACCA

>G1050 Amino Acid Sequence (domain in AA coordinates: 372-425)
MGGGGDTTDTNMMQVRVNSSSGTSSSSIPKHNHLNLPALIRSHHFRHPFTGAPPPPIPPPI
SPYSQIPATLQPRHSRMSQPSSFFSFDLPLNPSAPSVSVSVEEKTGAGFSPSLPPSP
FTMCHSSSSRNAGDGENLPPRKSHRRSNSDVTFGFSSMMSQNKSPPLSSLERSISGEDT
SDWSNLVKKEPREGFYKGRKPEVEAAMDVFTAYMNLNIDVLNSFGGEDGKNGENVEE
MESSRSGTKTKTNGSSSDSEGDSSASGNVKVALSSSSSGVKRRAGGDIAPTGRHYRSVS
MDSCFMGLNFGDESSLKLPPSSSAKVSPNTSGEGNSSAYSVEFGNSEFTAAEMKKIAD
EKLAEIVMADPKRVKRI LANRVSAARSKERKTRYMAELEHKVQTLQTEATTLQAQLTHLQ
RDSMGLTNQNSLKFRLQAMEQQAQLRDALSEKLNVEVQRLKLVIGEPNRRQSGSSSES
KMSLNPMEFQQLSISQLQHQMQHSNQCSMTMAKHTSND*

>G1463 (199..1209)
TATCCTTCGCAAGACCCCTTCTCTATATAAGGAAGTTTCATTTTCATTTGGAGAGGACACGC
TGACAAGCTGACTCTAGCAGATCTGGTACCGTCGACAGTTTGAGATTTGCTTCATCCGGT
TTTTTTATTTTCTGCAAAATATGTCACTCTCTCCATTTTGTTCATATATAATATGTTG
AAGTTTGATCAACTTAGTATGCGTTCTTTTCTCTCTAGTTCTCTGTTTCTTGGTCTGA
TTTAGTTTCGTTATGCGGACACACTGCTCAACGCAGAGACGAAGTAATAATCTCACGT
TATCTGAAGCCTATGATCGTTAACAGAGTATCATGGCCTGATCTCTTCATCGAAGACGCA

GACGTGTTCAACAAGGATCCATATGTGAAGTTCATGCTGAGATCCCTAGCTTCGTGATC
GTTAAACCACGAACAAAGGCTTGTGGTAAAACCGATGGATGTGATTCCGGTGTGCGGAGG
ATCATTGGTCGTGATAAGCTGATAAAGTCGGAGGAGACTGGTAAGATTCTAGGGTTCAAG
AAGATACTCAAGTTCTGCCTAAAGTGGAAACCTAGAGAATACAAGAGAAGTTTGGTAATG
GAAGAGTATAGGCTTACCAATAACTTCAACTGGAAGCAAGATCATGTGATTGCAAGATT
CGGCTTTTGTGTTGAAGCAGAAATTAGTTTCTTGCTAGCCAAGCATTCTTACACTACATCA
GACTCACTTCCTCGAAATGTGCTGTTGCCAGCTTATGGATTCTGTTTACCAGATAAACAA
GAGGAGGACGAATTTATCCGGTGACGATAATGATTTTCAAGAGAAAAGATTGGCCTAGC
TACGTTACCAACAACGTGTATTGTCTGCATCCATCGGAGCTTGTGAATGTTACGATGGG
AAGTTTCATGATAACGGAATCTGCATCTTCGCTAACAGGACTTGTGGTGTAAACCGATAAA
TGCAATGAAGGTTACTGGAAGATTAAGCACCGTGAGAAGCTGATCATGTACGGTACGGG
CAGACCATTGGTTGGAAGAAAGTTTTTTCAGTTTATGAAACGGAGAAAGAAAGACATTTT
GGTAATGGAGAAGAAGTGAAGGTAACCTGGACTCTAAAAGAGTATAGGCTTACCAGAAAA
ATGAACAAGAATAAAGTGGTGTGCGTTATCAAGTATAAGGTAAAGTGTTTACCGAGGATA
ACTAGCTAGGGACTTCTACTCTTGGTTTCATGATCGATCGCACCCTCTAGACAGGCCCTC
GTACCGGATCCTCTAGCTAGAGCTTTTCTGTTCTGATCATCGGTTTCGACAACGTTTCGTCA
>G1463 Amino Acid Sequence (conserved domain in AA coordinates:9-156)
MRFFSLVPLFLGRFSFVMADTLLNAEDEVIISRYLKPMIVNRVSWPDLFIEDADVFNKD
PYVKFHAIEIPSFVIVKPRTKACGKTGDCDSGCWRIIGRDKLIKSEETGKILGFKKILKFC
LKWKPREYKRSVMEEYRLTNFNWQDHVICKIRLLFEAEISFLLAKHFYTTSDSLPRN
VLLPAYGFCSPDKQEEDEFYPTIMISEGKDWPSYVTVNNVYCLHPSSELVNVHDGKFHDNG
ICIFANRTCGVTDKCNIEGYWKIKHREKLIMSRYGQTIGWKKVFQFYETEKERHFGNGEEV
KVTWTLKEYRLTRKMNKNKVVCVIKYKVKCLPRITS*
>G1944 (236..1306)
TCGACCTTCCTAATTTCCAACCTCTGTTCTTAGCAATATATTTTTTCTCCAAAAATAATT
CTCAGTTTGATTTTCTTCTTCTAGCTCTTAAGTATATTTCTTTGTTGTTATTTATCTTTT
AATCCTTTAATCTCATCTTTTATCTTTAATCAAAACCCAAAATTTACATGGGTTCTT
GAAAATCTAGAAGAAATAAAGGAAACATAACAAAAATAGAAAGAAAAAGAGCTAATGGT
CTTAAATATGGAGTCTACCGGAGAAGCTGTTAGATCAACCACCGGTAACGACGGTGGTAT
TACGGTGGTTAGATCCGACGCGCGCTCAGATTTCCACGTAGCTCAAAGATCAGAAAGCTC
AAACCAATCTCCACCTCTGTCACTCCTCCTCCACCACAGCCATCGTCTCATCACACAGC
TCCTCCGCGCTGCAAATTTGACGGGTGACGACTACGACTACGACGGCCCGATGGAAGG
TATCTCCGGTGGACTGATGAAGAAGAAGCGTGGACGGCCAAGGAAGTATGGACCGGACGG
GACTGTTGTAGCGTTATCTCTTAAACCGATTTCATCAGCGCCGGCGCGCTCGCATCTTCC
GCCGCGGAGTTACACAGTTCATCGATTTCTCCGCTTCTGAGAAACGTAGCAAAGTGAAACC
AACGAACCTGTTTAAACAGAACAAAGTATCATCACCAAGTTGAGAATTTGGGTGAATGGGC
TCCTTGCTCCGTCGGTGGTAATTTACACCTCATATAATCACAGTCAACACCGGCGAGGA
TGTAACAATGAAGATAATCTCGTTTTCGCAACAAGGACCTCGCTCTATTTGTGTTCTGTC
AGCAAACGGTGTATTTCAAGCGTTACACTTCGTGACCCAGATTCTCTGGCGGCACATT
GACATACGAAGGTCGGTTTGGATATTATCATTATCCGGGTCAATCATGCCTAATGATTC
AGGCGGAACACGAAGTAGAACGGGAGGAATGAGTGTATCGTTAGCAAGTCCCGATGGACG
TGATAGTAGGCGGTGGCCTCGCCGGTTTACTAGTAGCCGAGTCCGGTTTCAAGTGGTGTG
AGGAAGTTTTTTAGCGGGCACTGACCATCAAGATCAGAAACCGAAAAAGAACAAACATGA
TTTCATGTTGTGTCAGTCTTACCGCTGCAATTCCTATCTCTAGTGCAGCTGATCACCGGAC
AATCCATTCCGTCTCGTCTCTTCCGGTCAATAATAATACATGGCAGACTTCTTTAGCTTC
CGATCCAAGAAACAAGCATACCGATATTAATGTCAATGTAACCTGAAATCCAATCTTTCT
CTGTATTTTCTGTTAAACAAGTTTGATTTGGTTGTTTATCTACATAGGATTTTACTAAAA
TGGTAGTATTATTTATAGGGTTTTAGGGTCTTTATTTTGGTTCCACTGTTGTCACTTGT
GGATA
>G1944 Amino Acid Sequence (domain in AA coordinates:87-100)
MVLNMESTGEAVRSTTGNDGGITVVRSDAPSDHVAQRSESSNQSPSVTPPPPQPSHH
TAPPPLQISTVTTTTTTAAAMEGISGGLMKKKRGRPRKYGPDGTVVALSPKPISSAPAPSH
LPPPSHHVIDFSASEKRSKVKPTNSFNRTKYHHQVENLGEWAPCSVGGNFTPHIITVNTG
EDVTMKIISFSQQGPRISICVLSANGVISSVTLRQPDSSGGTLTYEGRFEILSLSGSFMPN
DSGGTRSRGTGMSVSLASPDGRVVGGLAGLLVAASPVQVVVGSFLAGTDHQDQKPKKNK
HDFMLSSPTAAIPISSAADHRTIHSVSSLPVNNNTWTQSLASDPRNKHTDINVNVT*
>G2383 (37..990)

GACCTCTTTGATCCCTTCATTCCCCATCAAACAACCATGTTTCCTTCTTTTCTTACTCAC
ATTCAAAGCCCTAATTCTCACCATCACTACTCTTCGCCTTCTTTTCTTTCTTCCGAT
TTTCTTGAGAGTTTTGATGAATCCTTCTTGATAAACCAATTCTTGTTACAGCAGCAAGAT
GTAGCAGCAAATGTTGTTGAATCTCCTTGGAATTTTGCAAGAAGCTTGAGCTTAAGAAG
AAGAATGAGAAGTGTGTTGATGGAAGCACCTCACAGAGGTTCAATGGAGAAGGACGGTC
AAAAAAGGGACAGGCATAGTAAGATCTGCACGGCTCAAGGTCCTAGAGACCGAGGATG
AGGCTGTCTCTTCAGATTGCTCGCAAGTTTTTCGATCTTCAAGACATGTTGGGTTTCGAC
AAGGCGAGCAAGACGATTGAATGGCTTTTCTCAAATCAAAGACTTCCATCAAACAACCTT
AAAGAAAGAGTGGCTGCATCGGAAGGAGGAGGAAAGGATGAACATCTCCAGGTTGATGAA
AAGGAAAAGGATGAGACACTGAAGTTGAGAGTCTCAAAGAGAAGAACAAGACTATGGAG
AGCTCTTTTAAGACTAAAGAGTCGAGAGAGAGAGCTAGAAAGCGAGCAAGAGAGAGAACA
ATGGCAAAGATGAAGATGAGATTATTTGAGACCTCGGAAACAATTTTCAATCCTCATCAA
GAACTAGAGAGATCAAGATAACCAATGGTGTACAATTACTAGAAAAGGAAAATAAAGAA
CAAGAATGGAGTAATACTAATGATGTTTACATGGTAGAGTATCAAATGGATTCTGTGAGC
ATCATAGAGAAGTTTCTTGGACTAACCAGTGACTCTAGCTCCTCTTCCATTTTGGTGAC
TCCGAGGAATGTTACACAAGTCTTAGTTTCAAGAGGTACAATTTTCAAGCAGCAGGTAAC
AGCAATGTGTTAACTAAAACCTTAATTGAGTAATGCAGTTTGTGATTAATATTAGCTTTT
TGGTAATTCCAGGAATGTCGACACCAAGGG

>G2383 Amino Acid Sequence (conserved domain in AA coordinates:89-149)

MFPSFITHIQSPNSHHYSSPSFPFSSDFLESFDESFLINQFLLQQQDVAAENVVESPWKF
CKKLELKKKNEKCVDGSTSQEVQWRRTVKKRDRHSKICTAQGPRDRMRRLSLQIARKFED
LQDMLGFPDKASKTIEWLFSKSKTSIKQLKERVAASEGGGKDEHLQVDEKEKDETLKLRVS
KRRTKTMESSEFKTKESRERARKRERERTMAKMKMRLFETSETISDPHQETREIKITNGVQ
LLEKENKEQWSENNDVHMVEYQMDSVSIIKFLGLTSDSSSSSIFGDSEECYTSLSVR
GTISAAGNSNVLTKNPN*

>G571 (326..1708)

TAGCCGACCTCTCTTCTCTTCTTGAAAAAACACCAAAGGAGCTTTAAATGCTCCGTTA
CATAATCTCTATCTCTTTCCAAGAATATAGAGAAAGGAAAATAATATACAAGAATTAAAA
GAAGGTATATCATCATCTCTCTAGCTAGTGATCAAAGCACCGTCATCATCATATATATC
ATCAGCTTGCCCTCAGAGGAGAAGACCAACATAAGAGAGATCGAAGATCAAAATCTATCTC
TCTTCATCATCTTCTGCTGTTACTATCATATCACACGCTCTCTCAAACATCATCCTATAT
ATAGACTTCTCTTCATCATCATCAAATGCAAGGTCATCACCAGAATCATCATCAACACTT
ATCATCATCTCCGCCACGCTCTTCCATGGAACTTCATGAACAAAGATGGGTATGATAT
TGGAGAGATAGACCCACTCTTCCCTCTATCTTGATGGACAAGGACATCATGATCCTCC
ATCAACTGCTCCTTCTCTTTACATCATCATCACACAACCTCAGAAATTTGGCGATGAGACC
TCCAACATCGACGCTCAACATCTTCCATCTCAGCCTATGCACATAGAGCCACCTCCTTC
TTCTACACACAATACCGATAATACAAGATTAGTTCGGGCTGCTCAACCTAGTGTTCCAC
TCGACCAGCTTCTGACCCGTCATGGACTTGACCAATCATTTCTCAGTTTCATCAACCTCC
TCAAGGTTCTAAATCCATCAAGAAGGAAGGGAACCGCAAGGGTCTTGCCCTCATCGGACCA
TGACATACCTAAATCGTCAGACCCTAAACATTGAGAAGACTAGCACAAAACAGAGAAGC
AGCAAGAAAAAGCAGATTACGTAAAAAGGCTTATGTTTCAAGCACTCGAGTCATGTAGGAT
CAAACCTGACCCAACTAGAACAAAGAGATTCAACGGGCCAGATCCCAAGGCGTATTCTTTGG
AGGGTCTCTTATAGGAGGAGATCAACAGCAAGGTGGACTACCCATTGGCCCTGGCAACAT
CAGCTCTGAAGCAGCGGTGTTGATATGGAATATGCGAGGTGGCTGGAGGAGCAGCAGAG
GCTATTAAACGAACCTAAGGGTGGCAACACAAGAACACTTGTCCGAGAACGAGCTTAGGAT
GTTTGTGGACACATGTTTAGCTCATTATGACCATTTGATTAACCTCAAGGCTATGGTCGC
TAAGACCGATGCTTCCACCTCATTTCTGGAGCATGGAAAACCTCCAGCTGAACGTTGCTT
CTTGTGGATGGGTGSTTTCCGTCCATCGGAGATCATTAAGGTGATTGTGAACCAGATAGA
ACCATTGACGGAGCAACAGATAGTTGGGATATGTGGGCTGCAACAGTCCACACAAGAGGC
CGAGGAGGCTCTCTCGAAGGCCTCGAGGCGTTGAATCAATCACCTTCCGATAGCATTGT
CTCTGACTCCCTCCCGCTGCTCCGACCACTTCCCTCTCATCTATCCAATTTTATGTC
ACACATGTCTTAGCTCTCAACAAGCTCTCTGCTCTCGAGGGCTTCGTTCTCCAGGCGGA
TAATTTGAGGACCAACGATCCATAGGCTGAACCAATTGTTGACGACCCGTCAGAAGC
ACGGTGTCTTCTAGCCGTTGCGGAGTACTTCCACCGTCTTCAAGCTCTAAGTTCTCTCTG
GCTAGCCCGTCTCTCGGCAAGATGGATAATACTAAAACAACCTGATGAAGGAAACCAAAAAC
AAAAACAAGAGAATAGGTTGATTAGTTAGCCGCCAGCTTGACCTCTTTATCATATATATC
GTCTCTCTACTCAAATACAGTGCAATTAGGGAATGTTTGGCTTCTTTTGGTATATG

ATTCTTACTATTATGTTTTTAATCAAGA

>G571 Amino Acid Sequence (domain in AA coordinates: 160-220)
MQGHHQNHQHQLSSSSATSSHGNFMNKDGYDIGEIDPSLFLYLDGQGHHDPPSTAPSPLH
HHHTTQNLAMRPPTSTLNIFFSQPMHIEPPPSSTHNTDNLRLVPAQAQPSGSTRPASDPSM
DLTNHSQFHQPQGSKSIIKEGNRKGLASSDHDIKSSDPKTLRRLAQNREARKSRLRK
KAYVQQLDESCRIKLTQLEQEIQRARSQGVFFGGSLIGGDQQQGGPLIGPGNISSEAAVFD
MEYARWLEEQQRLNLNLRVATQEHLSNELRMFVDTCLAHYDHLINLKAMVAKTDVFHLI
SGAWKTPAERCFLWMGGFRPSEIIKVIQVNIQIEPLTEQQIVGICGLQQSTQEAEEALSQGL
EALNQLSDSIVSDSLPPASAPLPPHLSNFMHMSLALNKLKSALEGFVLQADNLRHQTIH
RLNQLLTTRQEARCLLAVAIFYHRLQALSSLWLARPRQDG*

>G636 (6..1814)

CGATGATGCAACTGGGTGGTGGTACTCCGACCACTACAGCGCGGCTACAACCGTCACAA
CTGCTACAGCACCACCGCCACAATCAAACAACAACGATTACGCGGCAACAGAAGCAGCGG
CAGCAGCGGTTGGGGCGTTTGGAGGTGTCGGAAGAGATGCACGACCGTGGGTTTGGAGGAA
ATCGTTGGCCGCGGCAGGAAACGCTAGCGTTGTTGAAAATACGATCTGACATGGGAATAG
CGTTTCGAGACGCTAGCGTTAAAGTCCCTTATGGGAAGAGGTTTCTAGGAAAATGGCGG
AGCATGGTTACATAAGAAACGCAAGAAATGCAAAGAGAAATTCGAGAACGTTTACAAAT
ACCACAAACGAACCAAGAAGGTCGTACCGGAAAATCCGAAGGCAAACTTATCGCTTCT
TTGATCAATTAGAAGCTCTCGAGTCTCAATCTACAACCTCACTCCACCATCATCAACAAC
AAACGCCTCTTCGACCACAGCAAAACAACAACAACAACAACAACAACAACAACAGCT
CCATATTTTCAACTCCTCCTCCGGTAACGACAGTTATGCCGACGCTTCCTTCTTCATCAA
TTCCTCCGTATACTCAGCAGATTAATGTACCTTCGTTTCCAAACATCTCCGGTGATTTTC
TATCGGATAATTCTACATCGTCTTCGTCCTTCTTATTTCGACTTCTTCTGACATGGAGATGG
GTGGTGGAACTGCGACTACAAGGAAGAAAAGGAAGAGGAAATGGAAGGTGTTTTTCGAGC
GGTTGATGAAACAAGTAGTTGATAAACAGGAAGAGCTTCAACGCACATTCTTGGAAGCTG
TTGAAAAGCGAGAACACAAGAGATTGGTTAGAGAAGAGTCTTGAGAGATTCAAGAGATTG
CCAGAATCAACCGCAGCAGCAGATCTTAGCTCAAGAACGCTCTATGTCCGCTGCAAAAG
ACGCTGCTGTTATGGCCTTTCTTCAAAACTGTGAGAGAAACAACCGAATCAGCCACAAC
CGCAGCCTCAGCCGCAACAAGTTCGACCATCAATGCAGCTTAATAACAACAATCAGCAGC
AACCGCCTCAACGCTCTCCTCCACCGCAACCTCCTGCTCCGCTTCCGCGAGCCAATTCAAG
CGGTTGTGTCGACGTTAGACACAACGAAACGCACAATCGTGGTGATCAGAATATGACTC
CTGCAGCTTCAGCGAGCTCGTCCGCGTGGCCGAAAGTGGAGATAGAAGCATTGATAAAGC
TGAGGACGAATCTTGATTGCAAAATATCAAGAAAACGGACCAAAAGGACCATTGTGGGAAG
AGATATCAGCGGGAATGAGAAGGTTAGGATTCAACAGGAACCTCAAAGAGATGCAAGAGAA
AATGGGAAAACATAAAACAATACTTCAAGAAAGTCAAAGAGAGCAACAAGAAACGTCCCG
AAGATTCCAAGACTTGCCCTTACTTTACCAGCTTGATGCTTTATATAGAGAGAGGAACA
AATTCACAGCAACAACAACATTGCAGCTTCTTCTTCATCTTCCGGTCTTGTTAAACCGG
ATAATTCTGTTCCCTTGATGGTCCAACCAGAGCAGCAATGGCTCCGGCTGTAACGACTG
CGACAACTACTCCCGCAGCGGCTCAGCCTGATCAGCAATCTCAGCCGTCGGAGCAGAACT
TTGATGATGAAGAAGGTACAGATGAAGAGTACGACGATGAAGATGAGGAAGAGGAGAAATG
AAGAAGAGGAAGGAGGTGAGTTCGAGCTTGTGCCTAGCAATAACAACAACAACAAGACGA
CGAATAATCTGTAATGATGATGATTGAGTTCGAACCGGTTTGGTGGTGAAAGATTAGTA
ATCTTTTTTTAAGTTTTTGATACAGAACATGAGAATTTAAATATTGGAGGGTTT

>G636 Amino Acid Sequence (domain in AA coordinates: 55-145, 405-498)

MLGGGTPTTTAAATTVTTATAPPPQSNNDNDAATEAAAAAVGAFVSEEMHDRGFGGNNR
WPRQETLALLKIRSDMGIAFRDASVKGPLWEEVSRKMAEHGYIRNAKKCKEKFENVYKYH
KRTKEGRTGKSEGKTYRFFDQLEALESQSTTSLHHHQQTPLRPQQNNNNNNNNNNSSSI
FSTPPPVTVMPTLPLSSSIPPYTQQINVPSFPNISGDFLSDNSTSSSSSYSTSSDMMEMGG
GTATTRKKRKRKWKVFERLMKQVVDKQEELQRTFLEAVEKREHKRLVREESWRVQEIAR
INREHEILAQERSMSAAKDAAVMAFLQKLSKQPNQPPQPPQPPQVRPSMQLNNNNQQQP
PQRSPPPPAPLPPIQAVVSTLDTTKTHNRGDQNMTPAASASSSRWPKVEIEALIKLR
TNLDSKYQENGPKGPLWEEISAGMRRLGFNRNSKRCCKEKWENINKYFKVKESNKKRPED
SKTCPYFHLQDALYRERNKFHSNNNNIAASSSSGLVKPDNSVPLMVQPEQQWPPAVTTAT
TPAAAQPDQQSQPSEQNFDDEEGTDEEYDDEDEEEENEENEENEGGEFELVPSNNNNNNKTTN
NL*

>G878 (197..1738)

CAAAAAAATCTCTCCCATTAAGAACTGCCCAAAGAAATATTTTATACAAAATGAAAGA

GAGAAACACGACACGAATTTTGTATAATTAAGATTACACAAAAAAGTGTAGAAAGAG
 AAATATCTTCTTCTTTTTCTGTGTGAGTTGGGTTTGTAAAGTTTATCCTTTTTGTTC
 TCAAAATCAAGAATCGATGGCGGAGAAGGAAGAAAAAGAACCATCGAAGTTAAATCATC
 CACCGGAGTTTACGGCCAACGATTTCACTACCTCCTCGACCGTTTGGTGAAATGTTTTT
 TAGCGGTGGCGTTGGATTAGTCTTGGACCAATGACTCTCGTCTCAAATTTATTCTCTGA
 TCCTGATGAGTTCAAGTCTTCTCTCAGCTTTTAGCTGGAGCTATGGCTTCTCCGGCGGC
 AGCTGCTGTTGCCGCCGCTGCTGTGTTGCTACTGCTCATCATCAGACACCTGTGAGCTC
 TGTCGGTGATGGCGGTGGAAGCGGTGGTGATGTTGACCCGAGGTTAAGCAGAGTAGACC
 AACGGGATTGATGATAACTCAACCACCGGGGATGTTTACTGTACCGCCGGGGTTAAGTCC
 GGCTACTCTTTTGGATTCTCCGAGCTTCTTTGGTCTTTTTTACCTCTTCAGGGAACATT
 TGGTATGACACATCAACAAGCTTTAGCACAGTCACTGCACAAGCAGTTCAAGGCAATAA
 TGTTTCATATGCAGCAATCACAACAATCTGAATATCCTTCTTCTACACAACAACAACA
 ACAACAACAACAAGCTTCATTGACTGAGATTCCATCATTTTCTTCTGACCTAGGTCTCA
 GATTTCGAGCCTCGGTTCAGAAACATCGCAGGGTCAGAGAGAGACTTCGGAATATCTGT
 CTTTGAGCATCGGTTCACAGCCTCAAAATGCTGACAAACCAGCTGATGATGGATACAACTG
 GCGGAAATATGGGCGAAGCAAGTGAAGGGGAGCGATTTTCTCGGAGTTATTACAAATG
 TACGCATCCAGCTTGTCTGTCAAGAAGAAAGTGGAGAGGTCACTCGATGGACAAGTAAC
 GGAAATCATCTACAAGGGTCAACACAATCATGAGCTTCTCAAAGCGCGGTAACAATAA
 CGGGAGTTGTAAAGTTCTGATATTGCAAAATCAGTTTCAAACAAGTAATAGCAGTCTCAA
 CAAGAGTAAGAGGGACCAAGGAAACCAAGTTACAACAACAGAGCAGATGCTGAAGC
 AAGTGATAGCGAGGAGGTTGGGAATGCAGAGACTAGTGTGGGAGAAAGACATGAGGATGA
 GCCTGATCCCAAGCGAAGAAATACAGAAGTTTCGGGTTTCAAGAACAGTTGCTTCATCGCA
 TAGAATGTCAGAGAGCCTAGGATTATTGTCCAAACGACGAGTGAAGTTGACCTCTTAGA
 TGATGGATATAGGTGGCGCAAGTATGGTCAGAAAGTAGTCAAAGGAAATCCTTATCCGAG
 GAGCTACTATAAGTGTAACAACACAGATTGCGGAGTAAGGAAACATGTAGAGAGAGCAGC
 AACTGACCCAAAAGCTGTTGTAACAACATATGAAGGTAACATAACCATGATGTTCCAGC
 TGCTAGAACCAGCAGCCATCAGTTAAGACCAACCAATCAACACAACACCTCAACGGTTAA
 CTTCATCATCAACAGCCTGTTGCAGCTTTAAGGCTTAAAGAAGAGCAAAATCACTTGACA
 GAGAAGAAGAATACGACGGCGCTTGAGCTTTTGAGTGTAAATGAATCTTCTTTTGGTT
 AATGAACCTGTTTTTGTGCTTCAAAACACCAAGGTTTCTCTGGACAGAATCTCTGATA
 TTACAGTTTCAAAGGTATGTTCTTTTATTTCATGTTGGAATCTTCTGTGAATCTTAAG
 AAGCTTTAGGAGGTAAATGTAACCAACAGATTCAAAGTTATGCCCTTATGTGAATCTTT
 GTACATGGGATAAACAATAATTTACAGGTATCCTTTTTGTTCTTGTGTAAAAA
 AAAA

>G878 Amino Acid Sequence (domain in AA coordinates:250-305, 415-475)

MAEKEEKEPSKLSSTGVSRPTISLPPRPFGEFFSGGVGFSFGPMTLVSNLFSDFDEFK
 SFSQLLAGAMASPAAAAVAAAATAHQTVPVSSVGDGGSGGDDVDPFRKQSRPTGLMI
 TQPPGMFTVPPGLSPATLLDPSFFGLFSPLOQTFGMTHQQALAQVTAQAVQGNVHMQQ
 SQQSEYPSSTQQQQQQQQASLLEIPSFSSAPRSQIRASVQETSQSQRETSEISVFEHRS
 QPQNADKPADDGYNWRKYQKQVKGSDFPSYKCTHPACPVKKKVERSLDGQVTEIYK
 GQHNHELPOKRGNNNGSCKSSDIANQFQTSNSSLNKSQRDQETSQVTTTEQMSEASDSEE
 VGNAETSVGERHEDEPDPKRRNTEVRVSEPVASSHRTVTEPRIIVQTTSEVDLLDDGYRW
 RYKGQKVVGKNPYPRSYKCTTPDCGVRKHVERAATDPKAVVTTYEGKHNHDVPAARTSS
 HQLRPNNQHNTSTVFNHQPPVARLRLKEEQIT*

>G1134 (61..849)

TAAAGAAAGAGAAAAAAGCTTTCGTAGTGTCTATTGAAACCAGAGAAAAAGCCAAAGGGG
 ATGCAACCAACATCCGTCGGTAGTAGCGGCGGTGTTGACGACGAGGAGGCAGAGGAGGA
 GGAGGAGGGCTAAGTAGAAGTGGACTATCTCGGATCCGTTCACTCCAGCGACTTGGCTT
 GAAGCTTTACTTGAGGAAGATGAAGAAGAGTCTTTGAAACCTAATCTTGGTCTCACCAGT
 TTGCTTACCGGGAACTCGAACGATTTACCGACAAGTCGCGGCTCGTTTCGAGTTCCCGATT
 CCTGTTGAGCAAGGGTTGATCAACAAGGTGGGTTTCAACGACAGAATAGTACTCCGGCG
 GATTTTCTTAGTGGTTCTGATGGATTTATCCAAAGCTTTGGGATTCAGGCGAATTACGAT
 TACTTATCGGGGAATATCGATGTTTCTCCGGGAAGTAAGCGGTCTAGAGAAATGGAAGCA
 CTCTTCTCTTCTCTGAGTTTACTTCTCAATGAAAGGAGAGCAAAGCAGCGGTCAAGTT
 CCTACCGGAGTATCAAGCATGTGCGATATGAACATGGAGAACCCTTATGGAGGACTCTGTT
 GCTTTTAGGGTTCGGGCTAAACGTGGTTGCGCAACTCATCCCGCAGCATTGCCGAGAGG
 GTACGAAGGACGCGGATTAGTGATCGGATAAGGAAGCTACAAGAGCTTGATACCTAACATG

GACAAGCAAACCAACTGCAGACATGTTAGAAGAAGCAGTAGAATACGTGAAAGTTCTT
CAAAGGCAGATCCAGGAGTTAACAGAAGAACAGAAGAGGTGCACATGCATACCTAAGGAA
GAACAATAAGGTTTGCTCCTGATTTGTTTTATATTTGCTTAACGGCAATGATCTGATCGA
AAAATTCGAAAGATGATCTTAGCTTGAATTTAGATGGATGTCATGTTGAAAAGTATATTA
TTTGATAAATGGATGTAGGTGTAATATAAAATTTTGTACAATAATGAAGAAAGTTAAAA
AGAATTAATGAAAACATATATCTTTATGATATAAAAAAAAAAAAA

>G1134 Amino Acid Sequence (domain in AA coordinates: 198-247)

MQPTSVGSSGGGDDGGGRGGGGLSRSLRIRSRAPATWLEALLEEDEEESLKPNLGLTD
LLTGNSNDLPTSRGSFEFPIVEQGLYQGGFHRQNSTPADFLSGSDGFIQSFQIANYD
YLSGNIDVSPGSKRSREMEALFSSPEFTSQMKGEQSSQVPTGVSSMSDMNMENLMEDSV
AFRVRAKRGKATHPRISIAERVRRTRISDRIRKLQELVPMNDKQNTADMLEEAVEYVKVL
QRQIQELTEEQKRCTCIPKEEQ*

>G1008 (89..973)

GCCTTTTTGACTCTTCTTCTCTCTCTACTTTTTTTTCAGGCTCTCTCTCTATATCTCTA
TCTTCTTCTCCGGTTAACTAAAAGAGAAATGAAAAGCCGAGTGAGAAAATCCAAGTACAC
GGTTCACCGGAAAATCACATCCACACCGTTTCGACGGTTTCCCGAAGATTGTCAAAATCAT
AGTCACTGACCCATGCGCTACTGATTCTTCCAGCGATGAGGAAAACGACAACAAATCTGT
TGCTCCGAGGGTGAAAACGTTATGTGGATGAGATCAGGTTCTGTGACGAAGATGACGAACC
TAAACCGGCGAGGAAAACGAAGAAAAGTCCCCGGCGGCTGCGCGGAGAACGGTGGAGA
TTTGGTAAAGTCTGTGGTGAAGTATAGAGGAGTGAGACAACGACCTTGGGGAAAATTTGC
GGCGGAGATTCTGTGATCCTTCGAGTCGTACTAGACTCTGGCTTGGGACTTTTGGCAGCGC
GGAGGAAGCTGCTATAGGTTACGATAGAGCCGCGATTCTGAATCAAAGGTCATAACGCTCA
GACGAATTTTCTCACTCCTCCTCTAGTCCGACGACTGAGGTGTTACCGGAAACTCCCGT
GATTGACCTTGAACCTGTCTCTGGTTGTGATTCTGGCGAGGGAATCGCAAATCAGTCTGTG
TTCTCCGACTTCTGTCTTCTCCGTTTAGTCACAACGACGAAACAGAGTACAGAACAGAGCC
AACGGAAGAACAAAATCCGTTTTTCTTGCCTGATTTGTTTTCGCTCCGAGATTATTTTTG
GGATTCCGAAATTACCCCTGACCCTTTGTTTCTCGACGAATTCACCAGTCCTTGTTACC
AAACATCAACAACAACAACACAGTGTGTGATAAGGATACGAATCTGTCTGATAGTTTTCC
GTTGGGAGTGATCGGAGATTTCAGCTCATGGGATGTTGATGAGTTTTTCCAAGATCATT
GTTGGATAAGTAATTTGATGAGTTCTTCCCCAGAATTTTTCTGGGTTTCTCTTTTGGTT
GTGTGAGTGAGATGAGTGGTTTGTATGACAACGACGGGGATGAATCTTAGCCGTCGGTTTT
CCATTTCTGTTGACGGCTCCGATCAGCGGAAGAAGCGCAACGGAGTTTTTATTTATCTGTT
TGAGAATTTTATAATTTAATTTGCGAGTAAATATAGTAATTAGTGTTAAGATTGTGAGAG
TTTAAGTTAATTAGGGAGGGGTTTTGAATATTGGGGATTTTGGGAGGTTTTGTTTGGTT
TCTCTCCAAGTCTGTCACTATGCAAGGAAGCAGTATAAAGACCGTATATATATTTTATTA
TTAATATTGATAAAAGTAAAAAAAAAAAAAAAAAAAA

>G1008 Amino Acid Sequence (domain in AA coordinates: 96-163)

MKSRVRKSKYTIVHRKITSTPFDGFPKIVKIIIVTDPCATDSSSDEENDNKSVAAPRVKRYVD
EIRFCDEDEPKPARKAKKKSAAAAENGDDLKSVVKYRGVRQRPWGKFAAEIRDPSR
TRLWLGTFAATAEEAIGYDRAAIRIKGHNAQTNFLTTPPPSPTEVLPEPVIDLETVSGC
DSARESQISLCSPTSVLRFSHNDETEYRTEPTTEQNPFLLPDLFRSGDYFDSEITPDPL
FLDEFHQSLLPNNNNNTVCDKDTNLSDFSPLGVIGDFSSWDVDEFFQDHLDDK*

>G1020 (132..689)

CTGTTTCAAGAAAGCTCCCCAAAAGGAGCGTTGCTTTACTCTCTCTATAAAAAGAGCTC
TTCTACTTCTTCTCGTTACCACAAAACCTTTTCACCGATCTTCTCGTTCCATTCTTCTTC
CTAATTACACCATGCCCAACATCACCATGGGTTTGAACCCGACCCGGTTGCTCCAACGA
ACCCGACTCATCATGAGAGTAATGCTGCCAAAGAGATTCTGTACAGAGGCGTTAGGAAAC
GTCCATGGGGAAGATACGCCGCTGAGATCCGAGATCCGGTTAAGAAAACCTGAGTCTGGC
TCGGTACGTTTCGACACCGCTCAGCAGGCGCGCGTGTACGACGCGAGCCGCGCGTACT
TTCGTGGTGTTAAGGCTAAGACCAATTTTCGGTGTTATCGTTGGTAGTAGTCTACTCAGA
GTAGCACCGTCTCGTCTCTCCACGGCGGCACGGTTTATAACACCTCCGCACCTCGAGC
TCAGCTTAGGCGGCGCGCGCGTGTCTCGTAAGATCCCGCTTGTGCATCCGGTTTACT
ACTATAACATGGCGACGTATCCAAAGATGACGACGTGTGGTGTCCAGAGCGAGTCTGAAA
CGTCTGTCGGTCTGTTGATTTGGAAGGTGGAGCTGGGAAGATATCTCCGCCGTTAGATCTGG
ATCTTAACCTTAGCTCTCCGGCGGAATAGGCCGTGAGTTTTTTTTTCTTATGTCTGTTTC
TTTAGACAAAAAAAATAACGTTTCTTTTTTTTTCTGCCTAAGAAAAAATATTATCCG
TTTTTTAGAAGAAAAAAAATAAAAAA

>G1020 Amino Acid Sequence (domain in AA coordinates:28-95)
 MPNITMGLKPDVPAPTNPHHESNAAKEIRYRGVRKRPWGRYAAEIRDPVKKTRVWLGT
 DTAQQAARAYDAAARDFRGVKAKTNFGVIVGSSPTQSSTVVDSPTAARFITPPHLELSLG
 GGGACRRKIPLVHPVYYNMATYPKMTTCGVQSESETSSVVDFFEGGAGKISPLDLNL
 APPAE*

>G1023 (252..1250)

TCGTCCTTCTTAATCGCTTTCTGCTCTGTTTTCTCGTTCATCAAGCTACATCTACTAGCT
 CTCTCAGTGATTGATTTCTCACAGTTTCATCGATTTCCATGCGTTTAAGACCTAAAAGGA
 CTTGTCTCGGGTAAAGGACTTTTCTGTTCTTGAGAGAGTTCATTTTGAGGCTTTCTG
 GGAATTTTGAGAGGTTTTTGGGTAAAGGGGTTTGGTTTGAATTTGCGACACCAAG
 TGTTTCGATAAAATGGCTGAACGAAAGAAACGCTCTTCTATTCAAACCAATAAACCAACA
 AAAAACCCTGAAGAAGAAACCTTTTCAGCTAAATCACCTCCCAGGTTTATCTGAAGATT
 TGAAGACTATGAGAAAACCTCGTTTCGTTGTGAATGATCCTTACGCTACTGACTACTCAT
 CAAGCGAAGAAGAAGAAAGGAGTCAGAGAAGGAAACGTTATGTCTGTGAGATCGATCTTC
 CTTTCGCTCAAGCTGCTACTCAAGCAGAATCTGAAAGCTCATATTGTGAGGAGAGTAACA
 ATAATGGTGTAAAGCAAGACTAAAATCTCAGCTGTAGCAAAAAGGTTTTACGCAGCAAAG
 CATCTCCGGTCGTTGGACGTTCTTCTACTACTGTCTCGAAGCCTGTTGGTGTAGGCAGA
 GGAAATGGGGTAAATGGGCTGCTGAGATTAGACATCCAATCACCAGTAAGAATTTGGT
 TGGGTACTTACGAGACGCTTGAACAAGCAGCTGATGCTTATGTACCAAGAAGCTTGAGT
 TTGATGCTCTGCTGCGACCACTTCTGCTGCTTCTCTGTTTGTCAAATGAGTCTGGTT
 CTATGATCTCAGCCTCAGGGTCAAGCATTGATCTTGACAAGAAGCTAGTTGATTCGACTC
 TTGATCAACAAGCTGGTGAATCGAAGAAAGCGAGTTTTGATTTGACTTTGCGAGATCTAC
 AGATTTCTGAAATGGGTTGCTTCATTGATGACTCATTATCCCAAATGCTTGTGAGCTTG
 ATTTTCTCTTAACAGAAGAGAACAACAACCAATGTTGGATGATTACTGTGGCATAGATG
 ATCTGGACATCATTTGGTCTTGAATGTGACGGTCCAAGCGAACTTCCAGACTATGATTTCT
 CAGATGTGGAGATCGATCTTGGTCTCATTTGAACCAACCATTGACAAGTATGCTTTCGTTG
 ATCATATCGCAACAACCTACTCCCACTCCTTAAATATCGCGTGCCCAATAAGTTTGCAGC
 TAGGTGTTATTATTAGCTATAGGAGCAACGTAAAAAGCTCGTTGTTACTCGGTTTTGTCT
 TAAGTTATTAAAGTATAGCAGAGGCAGTTAATCTCAAGGGAAGCAAAAACCTAAAGATA
 GAAGCAGATGCAGTTTTGTGTGTTGGTGTACTAAAGAAAGTTTTGTTGACATAATGGTT
 TTGATGTTGTGGAGAAGATAGAGAGGTGTGATCGAAATTGTAAATCTCAGGTGGTTTTTT
 TTGAAGGCAATTTGTTCTATTAGGGTTTTTTTCTATATGAGGATTGTCTTTGAAAAGC
 CTTTAGATGTTTTCTAATTCGTAAGCTCTCTCAATCTTTGTAAGTTTTCCTGTTGAGTT
 ATTGATACATATGTGAGACCTACTTTATTTGTTTGTGCTACATACATTGTGATGGTTT
 CGTCAAAAAAAA

>G1023 Amino Acid Sequence (conserved domain in AA coordinates:128-195)

MAERKKRSSIQTNKPNKPKMKPFQLNHLPLGSEDLTMRKLRVFNDDPYATDYSSSEE
 EERSQRRKRYVCEIDLPAQAATQAESESSYQESNNNGVSKTKISACSKVLRSKASPV
 VGRSSTTVSKPVGVQRKWKWAAEIRHPITKVRTWLGTYETLEQAADAYATKKLEFDAL
 AAATSAASSVLSNESGSMISAGSSIDLKLVDSLTLDQAGESKKASFDFDFADLQIPE
 MGCFFIDDSFIPNACELDFLLTEENNNQMLDDYCGIDDLDIIGLECDGPSELDPDYFSDVE
 IDLGLIGTTIDKYAFVDHIATTTPTPLNIACP*

>G1053 (38..538)

GAACTCTTACATACTCATATAAACCAAACTAAAACCATGATTCCGGCAGAAATCAACGG
 ATATTTCCAATATCTATACCGGAATACAACGTAATAAACATGCCTTCATCTCCAACCTC
 TTCTTTAACTACCTAAACGATTGATCATCAACAACAACAACTATTCTCATCATCAA
 CAGTCAAGATCTCATGATAAGCAACAACCTCAACTTCCGACGAAGATCATCATCAAAGCAT
 CATGGTACTCGACGAGAGGAAACAGAGAAGGATGCTTTGAAACAGAGAATCTGCAAGGAG
 GTCAAGGATGAGGAAACAGAGACATCTTGATGAACCTCTGCTCAGGTAATAAGGCTTCG
 CAACGAGAACAACCTGTCTTATCGATAAGCTGAACCGCGTATCGGAGACTCAAAATTGTGT
 ATTGAAGGAGAACTCTAACTCAAGAAGAAGCTTCTGATCTCCGACAGCTTGTGTGTA
 ACTGAAATCTAACAAGAACAACAATAAGTTTTCCAAGAGAGTTTGAAGATAATTAGTA
 TTACTCAA

>G1053 Amino Acid Sequence (domain in AA coordinates: 74-120)

MIPAEINGYFYLSPEYNVINMPSSPTSSLNYLNDLIINNYYSSSSNSQDLMISNNSTS
 DEDHHQSIMVLDERKQRRMLSNRESARRSRMRKQRLHDELWSQVIRLRNENNNCLIDKLN
 VSETQNCVLKENSCLKKEASDLRQLVCELKSNKNNNNNSFPREFEDN*

>G1137 (202..1248)

TACTTCAGACTTCTACTCAAACCAAGTCACGTAGTTGGTTGGTGACATTTTCGCTGCATTTT
TCAATCTGTGATTGTTTTTCGTTTCGTCTTTCTTTTACTATTTTCTCGAAAAGGACACAAG
AAGTATTGCATTCACTCAGTTGAGCAACTTAACAATCGTGTGTACTTTTTGAAGTTCCC
TTGAGCTAAACTGCTAAGAGCATGCCTCTGGATAAGAGGCAACGGGATTTGCCTCTGGGC
TTAAGTCTCAAGCTTGCTTCAAGGATATAGTAGGTTCGGTCTGTCCTTCCTAGAATTCCT
CTCCCTGAGCTTGGGAACTATATGCAAGCTAAGCTTCAGGCTCGCTGTTTGCAGCCACCA
CCATTCCAGTCTTTGCTGTGCAGTCATGATAAGGAGTCTTATGGAAAAAGATTCTCACGG
TCTGACATGCGGTCTTGGTGCCTGCTGCTACTACTACTACTCCACTTGGAGCATT
GAGTCTTCTCAGAAAAGACTTTTGATATTCGATCAGTCAGGAGACCAGACTCGTCTATTA
CAATGTCCATTTCTCTACGGTTTCCATCTCATGCGGCTGCAGAACCAAGTGAAACTCTCT
GAGTTACAAGGTATAGAGAAAGCTTTCAAAGAAGATGGTGAAGAGTTTCAACAAGAGTGAT
GGAACAGAGTCAGAAATGCATGAAGACACTGAGGAGATCAATGCATTGCTATATTAGAT
GATGATTATGATGATGATTGCGAGAGTGATGATGAAGTAATGAGCACTGGTCACTCTCCT
TATCCAAATGAAGGAGTTTGCAACAAAAGGGAATTAGAAGAAATCGATGGTCTCTGTAAA
AGGCAGAACTACTGGATAAGGTCAACAACATCAGCGACTTATCATCACTTGTGGGCACT
GAGAGCTCCACACAACCTCAATGGATCTTCTTTCTTAAGGACAAAAGCTCCCTGAATCA
AAAACCATATCGACCAAGAGGACACTGGTCTGGTCTGAGCAACGAGCAGTCGAAGAAA
GACAAGATCCGCACAGCTCTGAAAATACTCGAGAGCGTAGTCCCTGGTGCAAAAAGGAAAC
GAAGCGCTCTTACTTCTGGACGAAGCAATTGATTACCTAAAGTTGCTGAAACGAGACTTA
ATCTCCACAGAGGTTAAGAACCAGCTCCACCACTCACAAGTCACCAATCTTGTGCTT
AAAGAGACAACATGGGGAACAAGAAATCTGCAGACAGATAAGGCGTGAAAGATTCTGACG
AGTTAAAACGTGTGAAGTGGGTTTTTGGGTACGTATCCTTGCACCAGCTTT

>G1137 Amino Acid Sequence (domain in AA coordinates: 264-314)

MPLDKRQRLPLGLSPQACFKDIVGRSVLPRIPLPELGKLYAAKLQARCLQPPPFQSLLC
SHDKESYGRFSSRSDMRSWCAAATTTTTPLGALESSQKRLIFDQSGDQTRLLQCPFPLR
FPSHAAEPVKLSELQIEKAFKEDGEEFHKSDGTESEMHEDTEEINALLYSDDDYDDDC
ESDDEVMSTGHSPYPNEGVCNKRELEEIDGPCKRQKLLDKVNNISDLSSLVGTESSTQLN
GSSFLKDKKLPESKTISTKEDTSGSLSNEQSKDKIRTKILESVPVPGAKGNEALLLLD
EADYLLKRLDLISTEVKNQSSSTHKSPILLKETTWTGRNLQTDKA*

>G1181 (113..1012)

CTCGATCTTTTAACCCCATTTATTACATATTACTCCTTCTACATTATTCTTCTCTGCT
TTCGTGACTTTTACGGGGACACTTTTGTTTTATAACTTACGCTTAAATCCTATGAATTC
GCCGCGGTTGACGCAATGATTACCGGAGAATCATCGTCACAAAGATCTATCCCAACGCC
GTTTCTCACAAAAACGTTTAACTCCTGTTGAAGATAGTTCCATCGACGATGTTATCTCATG
GAACGAAGATGGTTCCTCTTTCATCGTATGGAATCCGACAGATTTGCTAAAGATTTGCT
TCCTAAACACTTCAAACACAACAATTTCTCTAGTTTCGTTTCGTGAGCTCAACACTTACGG
ATTCAAAAAAGTTGTACCGGATCGATGGGAGTTTTCAAACGATTTCTTTAAGAGAGGAGA
AAAACGCTCTTCCCGTGAGATCCAACGTCGGAATAACAACGACGCATCAAACAGTTGT
TGCTCCTTCGTCGGAACAACGAAACGACGATGGTTGTATCACCGTCAAATTCGGGGGA
AGATAATAATAATAATCAGGTGATGTCTTCGTCTCCGTCGTGCTGATTTGTCATCAAAC
GAAGACGACTGGGAATGGTGGTTTATCAGTGGAGTTATTGGAAGAGAACGAGAAGCTTCG
GAGTCAAAACATTCAGCTAAACCGTGAGCTTACTCAGATGAAATCTATCTGCGATAATAT
CTATAGTCTCATGTGCAATTACGTCCGATCTCAGCCCACTGATCGGAGTTATTCTCCCGG
AGGTAGTAGTAGTCAACCGATGGAGTTTTTACCGGCGAAGCGGTTTTTCGGAGATGGAGAT
TGAAGAAGAAGAAGAAGCGAGTCCGAGGTGTTTGGTGTTCGGATTGGGTTAAACCGGAC
GAGAAGTGAAGGTGTTTCAAGTGAAGACGACGCGGTGGTGGGGAAAATTCGGATGAGGA
GACGCCGTGGTTGAGACATTATAATCGAACCAATCAGAGAGTTTGTAAATAAAAACGAAC
GGTTTAGATTTGTGGTGTAGATATGTGCGCGAAGTAGACGATTACAGCTTTTTAAGACAA
GCAGAGACGCTGTCCCTCTGTTTCAAGAAGTTTCTGCAATCTTGACTTCTTCTTTTAAAC
ACTTGTGTTTTTTTATTATTAAATTAATAACAATAAATGTTCTTTTTTTCAGTTTTGTTTC
TTCAAAAATAGTTCCGCTGTTTCTAGACTTTCTTTTTTT

>G1181 Amino Acid Sequence (domain in AA coordinates: 24-114)

MNSPPVDAMITGESSQRSIPTPLTKTFNLVEDSSIDDVISWNEDGSSFIWNPDTFAK
DLLPKHFPHNNFSSFVRQLNTYGFKKVVPDRWEFSNDFFKRGEKRLRLREIQRKITTTHQ
TVVAPSSEQRNQTMVVSNSNGEDNNNNQVMSSSPSSWYCHQTKTTGNGLSVELLEENE
KLRSQNIQLNRELQMKSI CDNIYSLMSNYVGSQPTDRSYSPGGSSSQPMEFLPAKRFSE

MEIEEEEEASPRFLGVPIGLKRTRSEGVQVKTTAVVGENSDEETPWLRHYNRTNQRVCN*

>G1228 (63..1139)

GCATTATAATTACTCACTCATCTTCTTTTCATTACATTACATACCAAACAAGAGCTCTC
AAATGGAAAGGTTTCAAGGACACATCAACCCCTGTTTCTTCGATCGAAAACCGGATGTGA
GAAGCCTCGAGGTTTCAAGGATTTGCAGAGGCTCAAAGCTTTGCTTTCAAAGAAAAAGAGG
AAGAAAGCTTACAAGATACAGTTCCATTTCTACAGATGCTGCAAAGTGAAGACCCCTCAT
CGTTTTTTTCAATCAAAGAGCCAAACTTTCTGACGCTACTGTCTCTTCAAACCCCTCAAGG
AGCCTTGGGAACTCGAAAGATATCTTCACTTGAGGATTACAATTTCAATTCACCGGTCC
AATCTGAGACCAACCGCTTCATGGAAGGAGCCAATCAAGCTGTGTCAAGCCAAGAAATTC
CCTTTAGCCAAGCAAACATGACACTCCCTTCTTCTACCTCATCACCCTCAGTGCACATT
CAAGACGAAAGCGCAAAATCAACCCTTGCTGCCTCAAGAAATGACTAGAGAAAAGAGAA
AGAGGAGGAAAAACAAACCAAGTAAAAACAATGAAGAGATTGAGAATCAAAGATAAACC
ACATTGCTGTTGAACGAAACAGAAGACGTCAAATGAACGAACATATCAACTCTCTCCGGG
CCCTTCTCCACCTTCTTACATCCAACGAGGAGACCAAGCTTCCATAGTAGGAGGAGCAA
TAACTACGTGAAGGTCCTCGAGCAAATCATACAATCTCTCGAATCGCAAAGAGAACGC
AACAACAAAGTAACAGTGAGGTAGTAGAAAACGCACTTAATCATCTCTCAGGCATTTCGT
CGAACGACCTGTGGACAACCTTTGAAGATCAAACCTTGATCCCCAAAATCGAAGCTACAG
TGATACAAAACCATGTGAGCCTTAAAGTTCAATGTGAGAAGAAACAAGGACAACCTTCTCA
AAGGAATCATATCACTTGAAAAGCTTAAACTCACTGTTCTTCACTCTCAATATCACTACTT
CGTCTCATTCTCTGTTTCTTATTCTTCAACCTCAAGATGGAAGATGAGTGCGACTTAG
AGTCAGCCGACGAGATTACGGCGGCTGTTTCATCGGATTTTCGATATTCGACAATTTGAT
TAAACACATATAATTCCAAAAATATTAAACAGCTGACAAAATGGTATCTTTGCGGCC

>G1228 Amino Acid Sequence (domain in AA coordinates: 179-233)

MERFQGHINPCFFDRKPDVRSLEVQGFAEAQSFQFKEKEEESLQDTPVFLQMLQSEDPSS
FFSLKEPNFLTLTSLQLKEPWELERYLSLEDSQFHSPVQSETNRFMEGANQAVSSQEIP
FSQANMTLPSSTSSPLSAHSRRKRKINHLLPQEMTREKRKRRTKPSKNNEEIEENQRINH
IAVERNRRRQMNEHINSLRALLPPSYIQRGDQASIVGGAINYVKVLEQLIQSLESQKRTQ
QQNSEVVENALNHLSSISNDLWTTLEDQTCIPKIEATVIQNHVSLKVQCEKKQGGQLLK
GIISLEKLLTFLHLNITSSSHSSVSYSFNLKMEDECDLESADETAHVHRIFDIPTI*

>G1277 (51..512)

ATTCTAAAGTCTCTCTCGGAAAGTAAGAGACTCAACTTCCGAGCCGCCATGACGCGG
GAGTAGCAGTAAAGCTGACGTGGCAGTCAAATGAAGAGAGAAAGACCATTCAAAGGGA
TCAGAATGAGAAAATGGGGGAAATGGGTTGCGGAGATTGAGAACCCAAAGCGTTCAA
GACTTTGGCTCGGCTCTTACTCTACTCCCAGCGGCGGCGCGTGCATACGACACGGCTG
TCTTTTACCTCAGAGGACCAACTGCTACGCTCAACTTCCCGGAGCTTCTGCCGTGTACCT
CCGCCGAGGATATGTGACGCGCAACGATCAGGAAAAAGGCGACGGAGGTGGGAGCTCAAG
TAGATGCGATAGGGGCGACGGTGGTGCAGAACAAACAGCCGCCGCGTCTTTTAGTCAAA
AGCGTGACTTTGGCGGCGGGTTATTAGAGCTTGTGACTTGAACAAGTTACCTGACCCGG
AAAATCTCGATGATGATTTGGTGGGAAAATAGACTGAAAAATAATAATAAATATCTTAC
AATGGTGGCTGTAGCTATCGTACGCGGAATGCTTGGGCTTGTGTTATATGACTACGTGGT
TACGGAAGATTCTCTGTTTCGTCATTGTATTAAAAATTAATCCCAAGTCAAACATA
CTGTACATTATTCTTAATTTAGTATTTCTTATTAATATCTATCATTTGTTTGGTGAACA
CCAGAATATTAGACTATTAATGTAAACGAGTTTTTAAATTTTCGATCATAATAACACCAAG
CTAGTTAAAGGTTAATATCTTGTACGAAGTCTTGAGTAAGTTCAATTGTATATATATG
TAACGGAAGAGGTTCTGTTCCGGTCCCAAGTGAAGTGGATCAAAGGTGACTTCACATAAAA
AATAAAAAAAA

>G1277 Amino Acid Sequence (domain in AA coordinates: 18-85)

MDAGVAVKADVAVKMKRERPFKGIRMRKWGWVABIREPNKRSRLWLGSYSTPEAAARAY
DTAVFYLRGPTATLNFPELLPCTSAEDMSAATIRKKATEVGAQVDAIGATVVQNNKRRRV
FSQKRDFGGGLLELVDLNKLDPDENLDDDLVGK*

>G1309 (53..859)

CGTCGACCTCTTAATTAAGACGACTTGAGAGAGAAAGAAAGATACGTGGAAGATGACCAA
ATCTGGAGAGAGACCAAAACAGAGACAGAGGAAAGGTTATGGTCACCTGAAGAAGACCA
GAAGCTCAAGAGTTTCATCTCTCTCGTGGCCATGCTTGCTGGACCACTGTTCCCATCTT
AGCTGGATTGCAAAGGAATGGGAAAAGCTGCAGATTAAAGGTGGATTAATTACCTAAGACC
AGGACTAAAGAGGGGGTCTGTTTAGTGAAGAAGAAGAAGAGACCATCTTGACTTTACATT
TTCTTGGGTAACAAGTGGTCTCGGATTGCAAAATATTACCGGGAAGAACAGACAACGA

GATTAAGAACTATTGGCATTCTCTATCTGAAGAAGAGATGGCTCAAATCTCAACCACAACT
CAAAAGCCAAATATCAGACCTCACAGAATCTCCTTCTTCACTACTTTCTTGCGGGAAAAG
AAATCTGGAAACCGAAACCCTAGATCACGTGATCTCCTTCCAGAAATTTTCAGAGAATCC
AACTTCATCACCATCCAAAGAAAGCAACAACAACATGATCATGAACAACAGTAATAACTT
GCCTAAACTGTTCTTCTCTGAGTGGATCAGTTCTTCAAATCCACACATCGATTACTCCTC
TGCTTTTACAGATTCCAAGCACATTAATGAACTCAAGATCAAATCAATGAAGAGGAAGT
GATGATGATCAATAACAACAACACTCTTCACTTGAGGATGTCATGCTCCGTACAGATTT
TTTGACGCTGATCATGAATATGCAAATTATTATTCTTCTGGAGATTTCTTCATCAACAG
TGACCAAAATTATGTCTAAGAAGAGTGAATATGATCGTAAGAGGAACATAAGCTAGTTAC
TTGTGTTACAGC

>G1309 Amino Acid Sequence (domain in AA coordinates: 9-114)
MTKSGERPKQRQRKGLWSPEEDQKLKSFILSRGHACWTTVPILAGLQRNGKSCRLRWINY
LRPGLKRGFSFSEEEEEITLTLHSSLGNKWSRIAKYLPGRTDNEIKNYWHSYLKKRWLKSQ
PQLKSQISDLTESPSSLLSCGKRNLLETETLDHVISFQKFSENPTSSPSKESNNNMIMNNS
NNLPKLFSEWISSNPHIDYSSAFTDSKHINETQDQINEEBVMNNNNNYSSLEDVMLR
TDFLPDPHEYANYYSSGDFFINSDQNYV*

>G1314 (1..990)

ATGGGAAGAGCTCCGTGTTGCGACAAGACAAAAGTGAAGCGAGGGCCTTGGTCGCCTGAA
GAAGACTCTAACTTAGAGATTACATTGAAAAGTATGGTAATGGTGAAATTGGATCTCT
TTCCCTCTCAAAGCCGGTTTGAGGAGATGTGGGAAGAGTTGTAGACTGAGGTGGCTAAAC
TATTTGAGACCAAAACATAAAGCATGGTGACTTCTCTGAGGAAGAAGACAGGATCATTTTT
AGTCTCTTCGTGCCATAGGAAGCAGGTGGTCAATAATAGCAGCTCATCTACCGGGACGA
ACAGACAACGACATAAAAACTATTGGAACACAAAGCTAAGGAAGAACTCTTGTCTTCT
TCCTCTGATTTCATCATCATCAGCCATGGCTTCTCTTATCTAAACCTATTTCTCAGGAT
GTGAAAAGACCAACCTCACCAACAACAATCCCATCTTCTTCTTACAATCCGTATGCTGAA
AACCTTAATCAATACCCAACAAAATCCCTCATCTCCAGCATCAATGGCTTCGAAGCTGGT
GACAAACAGATAATTTCTATATTAACCTAATTTATCTCAAGATCTCTATCTCTCGGAC
AGCAACAACAACACCTCGAACGCAAATGGTTTCTTGTCTCAACCACAATATGTGTGATCAG
TACAAGAACCACACAGTTTCTTCTCAGACGTCAATGGGATAAGATCAGAGATTATGATG
AAGCAAGAAGAGATAATGATGATGATGATGATAGACCACCACATTGACCAGAGGACAAAA
GGGTACAATGGGGAATTCACACAAGGGTATTATAATTACTACAATGGGCATGGGGATTG
AAGCAAATGATTAGTGGAACAGGCACTAATTCTAACATAAACATGGGTGGTTCAGGTTCA
TCTTCTAGTTCGATAAGCAACCTAGCTGAGAACAAAAGCAGTGGTAGCCTCCTACTAGAA
TACAAATGCTTGCCCTATTTCTACTCCTAG

>G1314 Amino Acid Sequence (domain in AA coordinates: 14-116)
MGRAPCCDKTKVKRGPWSPEEDSKLRDYIEKYGNNGNWSFPLKAGLRRCGKSCRLRWLN
YLRPNIKHGDFFSEEDRIIFSLFAAIGSRWSIIAHLPGRTDNDIKNYWNTKLRKKLLSS
SSDSSSSAMASPYLNPISQDVKRPTSPPTIPSSSYNPNYAENPNQYPTKSLISSINGFEAG
DKQIISYINPNYPQDLYLSDSNNTSNANGFLLNHNMCQYKNHTSFSSDVNGIRSEIMM
KQEEIMMMMHIDHIDQRTKGYNGEFTQGYNYNYNGHDLKQMISSGTGNSNINMGGSGS
SSSSISNLAENKSSGSLLEKCLPYFYS*

>G1317 (1..849)

ATGGGAAGATCACCTTGTGTGATAAAAATGGAGTGAAGAAGGGACCATGGACTGCTGAG
GAGGATCAGAACTCATCGATTATATTCGATTTTCATGGTCTGGCAATTGGCGTACGCTC
CCCAAAAATGCTGGACTCCATAGATGTGGA AAAAGCTGCCGTCTTCGATGGACCAATTAT
CTAAGACCGGACATCAAGAGAGGAAGATTCTCGTTTCGAGGAAGAAGAACTATCATTCAG
CTACACAGTGTTATGGGAACAAGTGGTCAGCAATAGCCGCTCGTCTACCAGGGAGGACC
GATAACGAAATAAAAACCATTGGAACACTCACATCCGCAAGAGACTTGTAGGAGTGGT
ATCGACCTGTACTACTCTCCACGCCTTGATCTTCTTGATTGTCTCTCACTTTTGAGT
GCATTTTCAACGACCAAACTTTTCAGCAGTTGCAACACATGCGTCTTCTCTCTTAAT
CCTGATGTATTGAGGTTGGCCTCTCTACTACTGCCACTTCAAAACCTAATCCAGTTTAC
CCATCGAACCTCGACCAAAATCTTCAAACCTCAAATACATCATCAGAATCGTCTCAACCA
CAAGCTGAGACTAGTACAGTCCCAACAACTATGAACTTCATCATTGGAGCCTATGAAC
GCAAGACTCGACGACGTTGGTCTTGCAAGATGTATTACCACCTTTGTGAGAGAGTTTGAC
TTAGACTCGCTCATGTCAACGCCAATGTCTTCTCCACGACAAAATAGCATTGAAGCAGAA
ACCAACTCCAGCACTTTCTTCGACTTTGGAATTCCGGAAGATTTTCATCTTAGATGACTTT
ATGTTTTTAA

>G1317 Amino Acid Sequence (conserved domain in AA coordinates:13-118)

MGRSPCCDKNGVKKGPWTAEEEDQKLIDYIRFHGPGNWRTLPKNAGLHRCGKSCRLRWNTY
LRPDIKRGRFSFEEEEETIIQLHSVMGNKWSAIAARLPGRTDNEIKNHNWTHIRKRLVRS
IDPVTHSPRLDLLDLSSLLSALFNQPNFSAVATHASSLLNPDVLRASLLPLQNPVY
PSNLDQNLQTPNTSSESSQPQAEFTVPTNYETSSLEPMNARLDDVGLADVLPPLSESF
LDLSLMSTPMSSPRQNSIEAETNSSTFFDFGIPEDFILDDFMF*

>G1323 (49..870)

AAGAGGGAATCTCAAAAGTGTGTGTCTGTGAGAGAGGAGAGAGAATATGGGCAAAGGA
AGAGCACCATGTTGTGACAAAACCAAAGTGAAGAGAGGACCATGGAGCCATGATGAAGAC
TTGAAACTCATCTCTTTTCATTACAGAATGGTCATGAGAATTGGAGATCTCTCCCAAAG
CAAGCTGGATTGTTGAGGTGTGGCAAGAGTTGTCTGTCTGCGATGGATTAATTACCTCAGA
CCTGATGTGAAACGTGGCAATTTCAAGTGCAGAGGAAGAAGACACCATCATCAAACCTTAC
CAGAGCTTTGGTAACAAGTGGTCGAAGATTGCTTCTAAGCTGCCTGGAAGAACAGACAAT
GAGATCAAGAATGTGTGGCATAACATCTCAAGAAAAGATTGAGCTCGGAAACTAACCTT
AATGCCGATGAAGCGGGTTCAAAAGGTTCTTTGAATGAAGAAGAGAACTCTCAAGAGTCA
TCTCCAAATGCTTCAATGTCTTTTGCTGGTTCCAACATTTCAAGCAAAGACGATGATGCA
CAGATAAGTCAAATGTTTGTGACACATTCTAACTTATAGCGAGTTTACGGGGATGTTACAA
GAGGTAGACAAACCAGAGCTGCTGGAGATGCCTTTTGTATTAGATCCTGACATTTGGAGT
TTCATAGATGGTTCAAGACTCATTCCAACAACCAGAGAACAGAGCTCTTCAAGAGTCTGAA
GAAGATGAAGTTGATAAATGGTTTAAAGCACCTGGAAAGCGAACTCGGGTTAGAAGAAAAC
GATAACCAACAACAACAACAGCATAAACAGGGAACAGAAGATGAACATTATCATCACTC
TTGGAGAGTTACGAGCTCCTCATACTAATGAAGCCATAAAGCAAGTCAATTTTACCTT
GAAAATGGAATTTATAGCTAACTTATTGGCATTATTAGTATATAAGCAAGATCAGATAGG
CGCATGTAGTAGCAACGAAGAAACGTCGAATTGTAGACAAAATGTAGATATTACAGA
GTTGAAAGATTGTATTTTGCAAATGATTGCTTTGTAGTGAAATCAAGTTATCACAAAAA
AAAAAAA

>G1323 Amino Acid Sequence (domain in AA coordinates: 15-116)

MGKGRAPCCDKTKVKRGPWSHDEDLKLISFIHKNHENWRSPLPKQAGLLRCGKSCRLRWI
NYLRPDVKRGNFSAEEEDTIIKLHQSFGNKWSKIASKLPGRTDNEIKNVWHTLKKRLSS
ETNLNADEAGSKGSLNEEENSQESSPNASMSFAGSNISSKDDDAQISQMFHILTYSEFT
GMLQEVDPKPELLEMPFDLDPDIWSFIDGSDSFQQPENRALQESEEDEVKWKHLESELG
LBENDNQQQQHKQGTDEHSSSLLESYELLIH*

>G1332 (1..606)

ATGGAATGCAAAAGAGAAGAAGGGAAGTCTTACGTGAAGAGAGGGTGTGGAAACCAGAA
GAAGATATGATATTAATAAGCTATGTTGAGACTCATGGTGAAGGAACTGGGCAGACATT
TCTCGTAGATCCGGGTTGAAGAGAGGAGGAAAAAGCTGTAGGCTGAGATGGAAGAACTAT
CTAAGACCAAATATCAAAAGAGGAAGCATGTCAACCACAAGAACAAGACCTTATCATCCGC
ATGCATAAGCTTCTTGGAAACAGATGGTCGTTGATCGCTGGTCGCCTTCCAGGTCGTACT
GACAATGAAGTGAAGAACTACTGGAATACTCATTTGAACAAGAAACCTAATTCCCGAAAA
CAGAATGCACCTGAATCAATCGTCGGCGCCACTCCTTTCAACGATAAGCCAGTTATGTCT
ACAGAAGTGAAGAAGAAGCCATGGAGAAGGAGGAGAAGAGGAGCAATACCTGGATGGAG
GAGACCAACCCTTTGGCTATGACGTCCACGTAGGATCTCCCTTGCCACTTATTTCCAC
TACCCAGACAACACTCTCGTGTGTTGACCCATGTTTTCTTTTACCGATTCTTTCTCTCTG
CTTTAG

>G1332 Amino Acid Sequence (conserved domain in AA coordinates:13-116)

MECKREEGKSIVKRLWKPEEDMILKSYVETHGEGNWADISRRSGLKRGKSCRLRWKNY
LRPNIKRGSMSPOEQDLIIRMHKLGNRWSLIAGRLPGRTDNEVKNYWNTHLNKKPNSRK
QNAPESIVGATPFTDPKVMSTELRRSHGEGGEEESNTWMEETNHFYDVHVGSPLPLISH
YPDNTLVFDPFCFSFTDFFPLL*

>G1334 (76..885)

ATAGCTCCCAACTAATAGGAATCTCAAGCTTCTCACTCTCTCTTGTTTTCCATTGGACT
TTTGGAAACATAAGCTATGCAAACTGAGGAGCTTTTGTGCGCCACCACAGACTCCTTGGTGG
AATGCTTTTGGATCTCAGCCGTTGACTACAGAGAGCCTTTCCGGCGAAGCTTCTGATTCA
TTCACCGAGTTAAGGCAGTTACTACGGAGGCAGAACAAAGGTGTGGTGGATAAAACAACT
TCTACAACCTCTCTTCACTTTCTACCTGGTGGTGAAGAGATTCAAGAGATGTGCCAAAG
CCTCATGTTGCTTTGCGATGCAATCAGCTTGCTTCGAGTTTGGATTGCTCAGCCAATG
ATGTACACAAAGCATCCTCATGTTGAACAATACTATGGAGTTGTTTCAGCATACGGATCT

CAGAGGTCTTCGGGCCGAGTAATGATTCCACTGAAGATGGAGACAGAAGAAGATGGTACC
ATCTATGTGAACCTCAAAGCAGTACCATGGAATTATCAGGCGACGCCAGTCCCGAGCAAAG
GCTGAAAACTGAGTAGATGCCGTAAGCCATATATGCATCACTCACGCCATCTCCATGCT
ATGCGCCGTCCTAGAGGATCTGGCGGGCGTTTCTTGAACACCAAGACAGCTGATGCGGCT
AAGCAGTCTAAGCCGAGTAATTCTCAGAGTTCTGAAGTCTTTCATCCGGAAAATGAGACC
ATAAACTCATCGAGGGAAGCAAATGAGTCAAATCTCTCGGATTCTGCAGTTACAAGTATG
GATTACTTTTCTAAGTTCGTCGGCTTATTCTCCTGGTGGCATGGTCATGCCTATCAAGTGG
AATGCAGCAGCAATGGATATTGGCTGCTGCAAACCTTAATATATGATCAGCAGATAGGGGA
CAAGACATGATTGGTCACCAGTCCTTTGTCTTGTCCCTTATCTTTCAGCCAAACGGAAA
GAGAACTTGTGTCTTGGAAAAAAGACATTGAGTTTCTTGGTTTATAAGATTGGTCCTTT
TACCATCCGTTTGGCTGTAAACAGGCAAATCATCTTGGCTCATGCTTCATCAAGTTCTT
ATCTTCGTCGTCTTCTTCTACGCATCTTCATAAGATCTCTGAACTAGTGAATAACATTT
CCTAGCATCATGTTTCACTAGTGTGTGTGTGAAGAACTCTGCCCTATTTCAGATGAT
GTATTGTGTGTAACTGTGTTTATGAAACAAACGTAAGACTTTCAAGTTAAAAA
AAAAAAAAAAAAA

>G1334 Amino Acid Sequence (domain in AA coordinates: 18-190)

MQTELLSPQTPWNNAFGSQPLTTESLSGEASDSFTGVKAVTTEAEQGVVDKQSTTLF
TFSPGGEKSSRDVPKPHVAFAMQSACFEFGFAQPMYTKHPHVEQYGVVSAYGSQRSSG
RVMIPKMETEEDGTIYVNSKQYHGIIRRRQSRKAELSRCKPYMHHSRHLHAMRRPR
GSGGRFLNKTADAQSKSPNSQSSEVFHPENETINSSREANESNLSDSAVTSMDYFLS
SSAYSPGGMVMPKWNAAAMDIGCCCKLNI*

>G1381 (32..802)

CAGCTTTAACAATACTCTCTCTCTCTCAAATGGGAAAACAAATCAACATAGAGAGTAG
TGCTACTCATCATCAAGACAATATTGTTTCCGTTATAACAGCCACGATATCCTCCTCCTC
CGTCGTAACGCTCTTCGTCAGACTCTTGGTCTACCTCCAAAGATCGTTAGTGCAAGACAA
TGACTCCGGAGGGAACCGCGGAAGAGCAACGTTAGTGATGATAACAAGAATCCGACGTC
GTATAGAGGAGTGAGGATGAGGAGTTGGGGAATGGGTGTCGGAGATTAGAGAGCCGAG
GAAGAAATCAAGAATATGGCTTGGCACTTATCCAACGGCAGAGATGGCAGCTCGTGCTCA
TGATGTGGCGGCTTTAGCTATTAAAGGCAACTCCGGTTTTCTTAATTTCCCTGAATTATC
CGGTTTGCTTCTCGTCCGTTAGCTGCTCTCCTAAGGATATACAAGCTGCAGCTACCAA
AGCCGCCGAAGCAACACGCTGGCACAAACCGGTTATCGATAAGAAATTAGCTGATGAGCT
AAGCCACTCTGAGTTGTTGTCTACCGCTCAGTCTTCGACTTCTAGTAGTTTCGTGTTTTTC
TTCGGACACGTCGGAGACTTCTAGTACGGACAAGGAAAGCAACGAAGAGACGGTGTTTGA
TTTGCCGGACCTTTTCACGGACGGGCTTATGAACCCAAACGATGCGTTTTGTTTATGCAA
CGGCACCTTTACGTGGCAGCTTTACGGAGAGGAGGATGTAGGGTTCAGGTTTGAAGAGCC
GTTTAATTGGCAAAATGACTAAACCGCCCTCCACTTGCTTACTGTAATTACTAACATATA
ATTTTCTTGATAAAGACATATATTTCCATTACGGTATTAACATACTTTTCTATCCTTT
TCTCTTTTCTTGTTTCTACATCTGAGTATATTGTCACTATGTGAAAAAATTGATCTCGTT
TTGAATATTTACTTTTCAAATTGAAGTAACGCAAGTGATGATAAAAAA
TTTGAATATTTACTTTTCAAATTGAAGTAACGCAAGTGATGATAAAAAA

>G1381 Amino Acid Sequence (domain in AA coordinates: TBD)

MGKQINIESSATHHQDNIVSVITATISSSSVVTSSSDSWSTSKRSLVQDNDSGGKRRKSN
VSDDNKNPTSYRGVVRMRWSGKWVSEIREPRKKSRIWLGTPTAEMAARAHDAALAIKGN
SGFLNFPESLGLLPRPVSCSPKDIQAAATKAAEATTWHKPVIDKKLADELHSELLSTAQ
SSTSSSFVFSSTSETSSDKBSNEETVFDLPDLFTDGLMNPNDAFCLCNGTFTWQLYGE
EDVGFRFEFPFNWQND*

>G1382 (90..1763)

CTCTCATTTTCGCCATAGCTGAGAGCTTCTTCTACTTTCCCTTAGCTTCTTTTTTCTTCA
TTTTTGTCTTACCCTGTGCGAATCTCTGAAATGAACCTCAAGCTAATGACCGGAAGGAGT
TTCAGGGAGATTGTTTCGGCGACGGGAGATCTCACGGCAAAGCACGATTAGCTGGAGGAA
ACGGAGGTGGAGGTGCTAGGTATAAGCTGATGTACCGGCCAAGCTCCGATCTCGAGGT
CGACTGATATCACGATTCCTCCTGGGTTGAGTCCGACTTCGTTTTTGAATCTCCTGTTT
TCATCTCCAACATCAAGCCAGAACCTTCCCTACTACTGGTTCTTTGTTCAAGCCTCGAC
CAGTGCACATTTCTGCTAGCTCAAGTTCTTATACAGGCAGGGGGTTCCATCAGAACACCT
TTACTGAGCAGAAGTCCAGTGAATTTGAGTTGAGACCTCCTGCATCAAATATGGTATATG
CAGAGCTTGGCAAGATTAGAAGTGAGCCACCAGTACATTTTCAAGGCCAGGGCCATGGAT
CCTCACACTCACCTTCTTCGATCAGTGATGCTGCAGGTTCCCTCAAGTGAGCTAAGCCGGC
CAACTCCTCCTTGTGATGACACCAACGAGCTCAGATATTCCGGCTGGATCTGATCAAG

AGGAATCAATCCAGACTTCCCAAAATGACTCCAGAGGAAGCACTCCATCCATCTTGGCTG
ATGATGGTTATAACTGGAGAAAATATGGTCAAAGCATGTCAAAGGGAGTGAATTTCCCC
GGAGCTATTATAAATGTACACATCCTAATTGTGAAGTGAAAAAGTTATTTGAAAGATCTC
ATGATGGGCAGATACCGGATATTATATACAAGGGTACACATGACCATCCTAAACCTCAAC
CTGGTCCGCCGAAACTCTGGTGGTATGGCTGCACAAGAAGAAAGGCTAGACAAGTATCCTT
CTTCAACTGGCCGAGATGAGAAGGGATCTGGCGTCTACAACCTTGTCTAACCCCAATGAAC
AACTGGTAACCTGAAGTACCTCCTATCTCAGCATCTGACGATGGTGGAGAAGCGGCAG
CGTCAAATAGGAATAAAGATGAGCCGGACGATGATGATCCATTCTCAAACCGGAGGAGGA
TGGAGGGTGCATGGAATAAATCCACTAGTGAAACCCATCCGGGAGCCTCGGGTTGTTG
TTCAAACCTCTGAGTGAGGTTGACATTCTGGATGATGGTTATAGATGGCGCAAATATGGGC
AGAAAGTCGTAAGGGGAACCCAAATCCCAGGAGCTACTACAAATGCACAGCTCATGGAT
GCCCAGTGAGAAAACAGTGGAGAGAGCATCACATGATCCAAAGCTGTAATAACAACAT
ACGAAGGCAAAACAGATCATGATGTTCCCACTTCAAAGTCTAGCAGCAATCAGAAATCC
AGCCTCGGTTCCAGACCAGATGAAACAGACACCATCAGCCTCAATCTTGGTGTGGAACTCT
CATCTGATGGACCTAACCACGCTTCCAACGAACATCAGCACCAGAATCAACAACCTTGTC
ACCAAACCTACCCAAATGGAGTCAATTTTCTAGGTTTGTTCATGCTAGTCCCATGTCATCCT
ACTATGCTAGCTTAAATAGCGGTATGAATCAGTACGGCCAGAGAGAAACAAAGAACGAGA
CTCAAATGGTGACATCTCGTCCTTGAACAATTCATCTTACCCATATCCGCCCAACATGG
GGAGAGTACAATCGGGTCCGTAAACAAAGTAAGCAACATTATGTACGGGATCTTCTT
AGGTTAGGAATGGGACGAGGCTTGTCTATATAATTCCTATTTCTTCACAGAGAGCTGA
TCTTGATTCAAACATATCTCCACATATATTTGTTGTGTACCTGTATTGAGTTCCAA
AAATGTTATGTAAAAATACACAAGATGTTAATGCTTTTATTTAAACAAGAAACAGCA
ATATTACTACAAAAA

>G1382 Amino Acid Sequence (domain in AA coordinates: 210-266, 385-437)

MNPQANDRKEFGQDCSATGDLTAKHDSAGNGGGGARYKLMSPAKLPISRSTDITIPPL
SPTSFLESPVFISNIKPEPSPTTGSFLKPRPVHISASSSSYTGRGFHQNTFTEQKSSEFE
FRPPASNVMVYAEGLKIRSEPPVHFQGGHSSSHSPSSISDAAGSSSELSRPTPPCQMTPT
SSDIPAGSDQEEISQTSQNSDRGSTPSILADDGYNWRKYGQKHVKGSEFPRSYYKCTHPN
CEVKKLFERSHDGQITDIIYKGTGHDHPKPPQRRNSGGMAAQEERLDKYPSSTGRDEKGS
GVYNLSNPNFNEQTNPEVPPISASDDGGEAAASNRNKDEPDDDDPFSKRRRMEGAMETPL
VKPIREPRVVVQTLSEVDILDDGYRWRKYGQKVVRGNPNPRSYYKCTAHGCPVRKHVERA
SHDPKAVITTYEGKHDHDPVTSKSSSNHEIQPRFRPDETDTISLNLGVGISSDGNHASN
EHQHQNQQLVNQTHPNGVNFVHASPMSSYYASLNSGMNQYQRETNETQNGDISSLN
NSSYPYPNMRVQSGP*

>G1435 (8..904)

GTGAAACATGGGGAAGGAAGTTATGGTGAGCGATTACGGTGACGACGACGGAGAAGACGC
CGGCGCGCGGATGAATATAGGATTCGGGAATGGGAAATTGGTTTACCCAACGGAGATGA
TTTGACTCCGTTATCTCAATATCTAGTCCCGTCGATTCTCGCGTTAGCTTTTACGATGAT
CCCAGAACGAAGCCGTACAATTCACGACGTCAATCGCGCGTCGAAATCAGCCTCTCTTC
GTTGAGAAGCAGTACCAATGCTTCTGTGATGGAGGAGGTCGTGGATCGAGTTGAATC
GAGTGTTCAGGATCAGATCCGAAGAAACAGAAGAAATCGGATGGTGGTGAAGCAGCGGC
GGTGGAGGATTCCACGGCGGAGGAAGGAGACTCCGGGCGTGAAGACGCGTCTGGGAAGAC
ATCGAAACGACCGCGTTTAGTGTGGACACCGCAGCTACACAAGAGATTTGTGGACGTTGT
GGCTCATCTAGGGATTAAAAACGCAGTGCCGAAGACGATTATGCAGCTGATGAACGTGGA
AGGACTTACTCGTGAGAACGTTGCGTCTCATTTGCAGAAATATAGGCTTTACCTTAAACG
GATTCAGGATTGACGACGGAAGAAGATCCTTATTCGTCGTCGGATCAGCTCTTCTCTTC
AACGCCGGTTCTCTCCACAGAGCTTTCAAGACGGCGGAGGAAGTAACGGAAAGTTGGGGGT
TCCGGTTCCGGTTCGTCGATGGTGCCTATTCCAGGCTATGGGAATCAAATGGGTATGCA
AGGATATTATCAACAGTATAGTAACCATGGCAATGAATCAAACCAATATATGATGCAGCA
GAATAAGTTTGGAAACAAATGGTGACATATCCTTCTGTTGGTGGTGGTGAAGTGAACAA
GTAAATGGATCTTAAAGGTCTATAATTTGCTCTACAGAGAGATACTGGTCTTGGCTTAT
GGTTTATTTTCCCACTTCATGAGGTTGTTGTGACTTTTAAATCTCCATGTTTTCCACACA
AGTCTTTATTTGCTTTGTATAGAAAATGATTTTCAGAAAATCACTGGGAAGCTTGGTATT
GTTGGAGGATGAAGCCTTCTATGAATGATTTAGTTTCTTACTGTCTCCATTCTTTATGAG
GTAATAAAGCCTTCTTTGCTCATCGCTTGTAGTCTTCTTAAATCAAGACAGCGTCACA
TGTTTGTTCGGTTATGTTAATGTTTCTTTCTTTGGATAATGAAGATAGCATCAGGTCTC
ATGTCTCCTCACTTTGATAAA

>G1435 Amino Acid Sequence (domain in AA coordinates: 146-194)
MGKEVMVSDYGGDDGEDAGGGDEYRIPEWEIGLPLNGDDLTPLSQYLVPSILALAFSMIPE
RSRTIHDVNRASQITLSSLRSSSTNASSVMEVVDRVSVPGSDPKKQKSDGGEAAAVE
DSTAEEDSGPEDASGKTSKRPLVWTPQLHKRFVDVVAHLGIKNAVPKTIMQLMNVEGL
TRENVAHLQKYRLYLKRIQGLTTEEDPYSSSDQLFSSTPVPQSFQDGGGSGNKLGVVPV
PVPSMVIPIPGYGNQMGMQGYQYQYSNHGNEQNQYMMQNKFGTMVITYPSVGGGDVNDK*

>G1537 (1..783)
ATGGAACAAACGAAGTAAACGCAGGAACAGCAAGCAGTTCAAGATGGAACCCAACGAAAGAT
CAGATCACGCTACTGGAATCTTTTACAAGGAAGGAATACGAACTCCGAGCGCCGATCAG
ATTCAGCAGATCACCGGTAGGCTTCGTGCGTACGGCCATATCGAAGGTAAAAACGTCCTT
TACTGGTTCCAGAACATAAGGCTAGGCAACGCCAAAAGCAGAAAACAGGAGCGCATGGCT
TACTTCAATCGCCTCCTCCACAAAACCTCCCGTTTCTTCTACCCCCCTCCTTGCTCAAAC
GTGGGTTGTGTCACTCCGTACTATTTACAGCAAGCAAGTGATCATCATATGAATCAACAT
GGAAGTGTATACACAAACGATCTTCTTACAGAAACAATGTGATGATTCCAAGTGGTGGC
TACGAGAAACGGACAGTCACACAAACATCAGAAACAACCTTTCAGACATAAGAACAACAGCA
GCCACAAGAATGCCAATTTCTCCGAGTTCAGTTCAGATTTGACAGATTTGCCCTCCGTGAT
AACTGTTATGCCGGTGAGGACATTAACGTCAATTCCAGTGGACGGAACAACTCCCTCTT
TTTCTCTTTCAGCCTTTGAATGCAAGTAATGCTGATGGTATGGGAAGTCCAGTTTTGCC
CTTGATAGTGATTCTCCGGTGGATTGTTCTAGCGATGGAGCCGGCCGAGAGCAGCCGTTT
ATTGATTTCTTTCTGTTGTTCTACTTCTACTCGTTTTGATAGTAATGGTAATGGGTTG
TAA

>G1537 Amino Acid Sequence (domain in AA coordinates: 14-74)
MENEVNAGTASSSRWNPTKDQITLLENLYKEGIRTPSADQIQQITGRLRAYGHIEGKNVF
YWFQNHKARQRQKQKQERMAYFNRLLLHKTSRFFYPPPCSNVGCVSPYYLQQASDHMNQNH
GSVYTNDDLHRNNVMIPSGGYEKRTVTQHQQQLSDIRTTAATRMPIPSSSLRFDRFALRD
NCYAGEDINVNSSGRKTLPLFPLQPLNASNADGMGSSSFALGSDSPVDCSSDGAGREQPF
IDFFSGGSTSTRFDSNGNGL*

>G1545 (67..729)
CATCACCAATCTTTTGAATCTAAGAGAGAGAAGAAGAAGGTCTAGAGAACGAAAAGA
AGAAACATGAATAACCAGAATGTAGATGATCATAATCTTCTACTCATTTCTCAATTGTAC
CCTAATGTCTATACTCCATTAGTACCACAACAAGGAGGAGAAGCAAAACCAACACGGCGG
AGGAAAAGGAAGAGCAAGAGTGTGTGGTGGCAGAGGAGGTGAAAACGAAGGCAATGGG
TGGTTTGAAGAGAGAAAATTTGAGTGATGAGCAAGTAAGAATGTTGGAGATTAGCTTTGAA
GACGATCATAAGCTTGAATCCGAGAGGAAAGATCGGCTTGCTTCTGAGTTAGGGCTTGAT
CCTCGTCAAGTCGCCGTCTGGTTCCAAAACCGCCGTGCACGGTGGAAAGAACAAACGAGTC
GAGGATGAATACACTAACTCAAGAATGCATACGAAACCACCGTCGTTGAGAAATGTTCGT
CTTGATTCTGAGGTTATTACCTAAAGGAACAACCTTACGAGGCTGAAAGAGAGATCCAA
CGGCTTGCAAAAAGAGTTGAAGGAACCTTTAAGTAACAGTCCTATCTCATCCTCTGTGACC
ATTGAAGCCAATCATACGACACCGTTTTTTGGAGATTACGACATCGGATTTGACGGTGAG
GCTGACGAGAACTTGCTCTACTCGCCAGATTACATTGATGGATTAGACTGGATGAGCCAA
TTTATGTAAAAAACTATAAGCTAATCTATTTTCAGTCGTAGTATAG

>G1545 Amino Acid Sequence (domain in AA coordinates: 54-117)
MNNQNVDDHNNLLISQLYPNVYTPLVPQQGGEAKPTRRRKRKSKSVVVAEEGENEGNGWF
RKRKLSDEQVRMLEISFEDDHKLESEKDRLASLGLDPRQVAVWFQNRRLRWKNKRVED
EYTKLKNAYETTVVEKCRLDSEVIHLKEQLYEAREIQRLAKRVEGTLSNSPISSSVTIE
ANHHTPPFGDYDIGFDGEADENLLYSPDYIDGLDWMSQFM*

>G1641 (1..867)
ATGGAGGTTATGAGACCGTCGACGTCACACGTGTCAGGTGGGAACCTGGCTCATGGAGGAA
ACTAAGAGCGCGCTCGCAGCTTCTGGTGAAGGTGCCACGTGGACGGCGGCAGAGAACAAG
GCATTGAGAAATGCTTTGGCGGTTTACGACGACAACACTCCTGATCGGTGGCAGAAGGTG
GCTGCGGTGATTCCGGGGAAGACAGTGAGTGACGTAATTAGACAGTATAACGATTTGGAA
GCTGATGTCAGCAGCATCGAGGCCGTTTAAATCCCGGTCCCCGGTTACATCACCTCGCCG
CCTTTCACTCTAGATTGGGCCGGCGCGGTGGCGGATGTAACGGGTTTAAACCGGTCAT
CAGGTTTGTAAATAACGGTCGCAGGCCGGTAGATCGCCGGAGCTGGAGCGGAAGAAAGGC
GTTCTTTGGACGGAGGAAGAACAAGCTATTTCTAATGGGTTTGAAGAAATATGGGAAA
GGAGATTGGAGAAACATATCTCGGAACCTTTGTGATAACGCGAACGCCAACACAAGTAGCT
AGCCACGCCCCAAAGTACTTCATCCGCAACTTTCCGGCGCAAGGACAAGAGACGAGCA

AGCATTACACGACATAACCACCGTAAATCTCGAAGAGGAGGCTTCTTTGGAGACCAATAAG
 AGCTCCATTGTTGTTGGAGATCAGCGTTCAAGGCTAACCGCGTTTCTTGGAAACCAAACG
 GACAACAATGGAACACAGGCAGACGCTTCAATATAACGATTGGAAACGCTATTAGTGGC
 GTTCATTACATACGCCAGGTTATGATTGGAGGGTATAACAATGCAGATTCTTGCTATGAC
 GCCCAAAACACAATGTTTCAACTATAG

>G1641 Amino Acid Sequence (domain in AA coordinates: 139-200)
 MEVMRPSTSHVSGGNWLMEETKSGVAASGEGATWTAENKAFENALAVYDDNTPDRWQKV
 AAVIPGKTVSDVIRQYNDLEADVSSIEAGLIPVPGYITSPPTLDWAGGGGGCNGFKPGH
 QVCNKRSQAGRSPELERKKGVPTWEEHKLFLMGLKKYKGDWRNISRNFVITRPTQVA
 SHAQKYFIRQLSGGKDKRRASIHDIITVNLEEEASLETNKSSIIVVDQRSRLTAFPNQT
 DNNGTQADAFNITIGNAISGVHSYQVMIGGYNNADSCYDAQNTMFQL*

>G165 (19..699)
 CTTCAAACATCTAAAAAATGGTGAAAAAACTCTTGGTCGTAGAAAGGTAGAGATAGTG
 AAAATGACTAAGGAATCAAACCTTCAAGTCACATTTTCCAAGAGAAAAGCTGGTCTTTT
 AAGAAGGCTAGTGAATTTTGCACATTATGTGATGCAAAAATTGCGATGATCGTGTTC
 CCAGCTGGAAAAGTATTTTCTTTTGGTCATCCAAATGTTGATGTTCTGCTTGACCACTT
 CGAGGGTGTGTTGTAGGACACAACAACACAAACCTTGATGAAAGCTACACAAAGCTTCAT
 GTTCAAATGCTCAACAAATCCTACACTGAGGTGAAGGCGGAAGTAGAAAAAGAACAAAAG
 AATAAGCAGTCGCGGGCTCAAATGAAAGAGAAAACGAAAACGCTGAGGAGTGGTGGAGT
 AAGTCTCCATTAGAACTCAACTTAAGTCAATCAACCTGTATGATACGTGTTCTTAAAGAT
 TTGAAGAAGATAGTTGATGAAAAAGCAATTCAATTAATCCATCAACAAACCCAACTTC
 TATGTTGGAAGTTCTAGCAATGCTGCTGCTCCAGCAACTGTTAGTGGTGGTAATATCTCC
 ACAACCAGGGGTTCTTGTATCAAAACGGAATGACGACTAATCCTACTCAAAACACTTCTG
 TTTGGATTTGATATTATGAATCGCACACCAGGAGTTTAAATAAGTCTATCCTCATTATGG
 GTCTTGGTACTATAAGTTCATCTCTCTCGTTGTTGACTTTTTTAAGTCTCCAATAGTTTGT
 TGTG

>G165 Amino Acid Sequence (conserved domain in AA coordinates: 7-62)
 MVKKTILGRRKVEIVKMTKESNLQVTFFSKRKAGLFKKASEFCTLCDAKIAMIVFSPAGKVF
 SFGHPNVDVLLDHFRCVVGHNNTNLDES YTKLHVQMLNKS YTEVKA EKEQKNKQSR
 QNERENENAE EWSKSPLELNLQSSTCMIRVLKDLKKIVDEKAIQLIHQTNPNFYVGS
 SSS
 NAAAPATVSGGNISTNQGFDFQNGMTTNPQTLLFGFDIMNRTPGV*

>G1652 (77..1078)
 AGCAAGTCCAAATCTCCCTCTCTCTCTCTATCTATCTCTCTATAGAAGATTTTTTAAC
 TAAGAAGCTAGCGATCATGGCCACAGCGATGAACGTTTTCTCTACCAAATGGTCTCCGA
 ATTGGATATAGAAGAATATAGTATCATCCACCAATCCACATGAACCTCACTCGTCGGAGA
 TGTTCCACAGTCTCTCTCATCTCTTGATGATACCACCACTTGTTATAACCTTGATGCTTC
 TTGTAATAAAAGTTTGGTGAAGAAAGACCTTCAAAGATCCTCAAGACCACTCACATATC
 ACCAAACTTACATCCTTTTTCTTCTTCTAATCCCTCCTCCAAAGCACCAGCCCTCTTC
 TAGGATCTTTCTTTTGAAGACAGGTTTACATGTTATGAATCACAACCTCTCCAACTT
 AATATTTAGCCCCAAGGACGAAGAAATTGGATTACCAGAGCATAAGAAAGCCGAGCTGAT
 AATAAGAGGGACAAAGAGAGCTCAATCCTTGACTCGAAGCCAATCAAATGCTCAAGATCA
 CATACTGGCAGAGAGAAAACGGAGAGAGAAGCTTACTCAAAGATTTGTAGCTCTTTCCGC
 GCTAATCTCTGGCCTAAAGAAGATGGACAAGGCTTCTGTGTTGGGAGATGCAATAAAGCA
 TATAAAGTACCTCCAAGAGAGTGTGAAAGAGTATGAGGAACAAAAGAAGGAAAAGACAAT
 GGAATCAGTGGTTCTTGTAAAGAAGTCTAGTCTGGTTTATAGATGAAAATCATCAACCATC
 ATCATCATCTTCTCAGATGGAATCGCAATAGCTCGAGCTCAAATCTCCAGAAATAGA
 AGTTAGGGTTTCAGGAAAAGATGTTCTTATTAAGATCCTATGCGAGAAGCAAAAGGGTAA
 TGTGATCAAGATTAAGGGGAGATTGAAAAGCTTGGTTTGTCTATCACCACAGCAATGT
 CTTGCCCTTTGGACCCACTTTTGACATCTCTATTATCGCTCAGAAGAATAACAATTTTGA
 TATGAAAATCGAGGATGTTGTGAAGAACTTGAGTTTGGCTTATCAAAGCTCACTTAATT
 GGTTCACGTTACATACATATACACATTATCATCGATTCTCCGATCGAAGAATCCAAA
 ATCAGTTTTTCCATGAAAGTGGTTTTTAGTTGTTAAGTTTGGTTGATGGAGATTCTTAA
 GTCATTTAAAGATCCTTGTCTTGTGTTGTTAAGTGTGCTTTAAGATGCATATCATCAA
 TGTTTTAGTAATTATTTCTCTCCAGTTTCATTTGGGACGGAATTTTTTTCGAGTTGTTGG
 ATATATATTTCTCGCATGTAAAGCATTTCTGTTAGTTTAAATAAACGTCCGATATGTTTCT
 TTGAAAA

>G1652 Amino Acid Sequence (domain in AA coordinates: 143-215)

MATAMNVFSTKWSELIDIEEYSIIHQFHMNSLVGDVPQSLSSLDLDTTTCYNLDASCNKSL
VEERPSKILKTTHISPNLHPFSSSNPPPKHQPSRILSFEKTGLHVMNHNSPNLIFSPK
DEEIGLPEHKKAELIIRGTRKRAQSLTRSQSNADHILAEKRREKLTQRFVALSALIPGL
KKMDKASVLGDAIKHIKYLQESVKEYEQQKEKTMESVVLVKSSSLVDENHQPSSSSSS
DGNRNSSSSNLPEIEVRVSGKDVLIKILCEKQKGNVIKIMGEIEKLGLSITNSNVLFPFGP
TFDISIIAQKNMNFDMKIEDVVKNSLSPGLSKLT*

>G1655 (132..755)

TTTCTAACTAGTCACATTGAGAGAGAGAGAGAGAGAAAGAGAGACTCTCAGAATCTGAAG
AAGAAGAAGAGATTGTTGTTTTTGCCTTTTATCATCGGTTTCTTTGAATCTCTGGTTTTA
AATCGGATTTAATGGTGGAGTCTCTGTTCCCGAGCATCGAAAACACAGGTGAATCGTCTC
GAAGAAAGAGCCGAGGATATCAGAGACGGCGGAGGCGAGATAGAGGCACGACGTGTCA
ACGAAGAAAGCTTGAAGAGATGGAAGAACGAATCGTGTGCAACAGATCTACGTTGTAAGC
TCGTGCAAGCTTTACGCCGAGTTCGTGAGAGATCTTCCACCACCAGCAACAACGAGACCG
ATAAATCGTCTCCGGCGCGGCGAGGGAGATACGTGATACGGCGGATCGAGTTCTAGCTG
CGTCCGCTCGTGGTACGACTCGGTGGAGCAGAGCGATTTAGCGAGTCGCGTCCGAGCGA
AGCTGAAGAAACATAGAAAGGCGAAAAAGTCAACGGGAAATTGTAAATCGAGAAAAGGTC
TCACGGAGACGAATCGGATTAAGTTACCGGCGGTTGAGAGAAAAGTGAAGATTCTTGGCC
GTTTGGTTCTGTTGTTGCCGAAAGTCTCTGTACCGAATCTTTTAGATGAAGCGACCGATT
ACATCGCAGCGTTAGAGATGCAGGTTGAGCCATGGAGGCTCTCGCCGAATTTTAACCG
CAGCCGACCCACGGACGACGTTGACCCGAACTTAACGGCGGCGAGTTAGTTTGTGAGTTGT
TAATTAGCTTTTCTTTTACCTTTTACCCCTTTATTTTGGCTTCAAGTGTTTTTTTTTTC
TCGTCGACGCGATTTTAATTTATTAATTC

>G1655 Amino Acid Sequence (domain in AA coordinates: 134-192)

MVESLFPSIENTGESSRRKKPRISETABAEIEARRVNEESLKRWKTNRVQIYACKLVEA
LRRVRQRSSTTSNNETDKLVSGAAREIRDTADRVLAASARGTTRWSRAILASRVRAKLLK
HRKAKKSTGNCKSRKGLTETNRIKLPAVERKLKILGRLVPGCRKVSVPNLLDEATDYIAA
LEMQVRAMEALAEELLTAAPRTTLTG*

>G1671 (188..751)

TCCCACTATCCTTCGCAAGACCCTTCCTCTATATAAGGAAGTTCATTTTCATTGGAGAGG
ACACGCTGACAAGCTGACTCTAGCAGATCTGGTACCGTCGACCCTCTCTATATAATCTTC
TTCTACACACACACACACGCAACCATATACGTACATGTGAAGTAGTGAGATCAATATC
GTTAGCAATGAATCTACCACCGGATTTAGGTTTTTTCCGACCGATGAAGAGCTCGTCGT
TCACTTCCTCCACCGAAAGCTTCCCTCTTGCCTTGTCACCCTGATGTGATCCCCGACCT
TGATCTTTACCATTACGATCCTTGGGACCTTCCCGGAAAGCTTTGGGAGAAGGGAGGCA
ATGGTACTTCTATAGTAGAAAGACACAAGAGAGAGTGACAAGCAATGGGTATTGGGGATC
AATGGGAATGGACGAGCCAATCTACACAAGCTCCACACACAAGAAAGTGGGAATCAAAAA
GTATCTAACTTTCTATCTCGGAGATTCTCAGACTAATTGGATCATGCAAGATATTCCCT
CCCGGATTCTCTTCTTCTATCTAGTCGATCTTCTAAGAGATCAAGCCGTGCTTCTAGTTC
TAGTCACAAACCCGATTATAGCAAGTGGGTGATATGCAGAGTGTATGAGCAAAATTGCAG
TGAGGAGGAAGAGGATGATGGGACAGAACTCTCATGTTTGGATGAAGTGTTTTTGTCTTT
AGATGATCTTGACGAAGTAAGCTTACCCTAATAAGACAGAAGCACCCCAAGAAGAGAAAA
AAAAAAAAGGGTTTAGTGGGCAATTATTTCTAAGCGACCGCTCTAGACAGGCCTAGTAC
CGGATCTCTAGCTAGAGCTTTCGTTCTGATCATCGGTTTCGACAACGTTCTGTCAGT

>G1671 Amino Acid Sequence (domain in AA coordinates: TBD)

MNLPPGFRFFPTDEELVVHFLHRKASLLPCHPDVLPDLPLYHYDPWDLPGKALGEGRQWY
FYSRKTQERVTSNGYWGSMGMDPIYTSSTHKKVGIKKYLTFFYLEGDSQTNWIMQEYSLPD
SSSSSSSRSSKRSSRSSSHKPDYSKWVICRVYEQNCSEEDDDGTELSCLDEVFLSLDD
LDEVSLP*

>G1756 (71..1003)

ATATGTACTTGTACACCAACCCACCAAAAGAGATAAAAGAGGAAACAAAAACTCGAAAAG
AGAGAGATATATGGGTGAGGTGGCTTATATGGACGAAGGAGACCTAGAAGCAATAGTCAG
AGGCTACTCCGGCTCCGAGACGCGTTTTCCGGCGAAAGTTCGGTACGTTTTACCTTC
GTTTTGCCTACCGATGGAGACGTCTAGTTTCTACGAACCGGAGATGGAGACAAGTGGCTT
AGATGAGCTCGGTGAACTTTACAAACCTTTTACCCTTTCTCCACACAACGATCCTCAC
AAGCTCGGTCTCTCTCCCTGAAGATTCAAAACCTTTCCGAGATGACAAGAAAACAACGATC
ACATGGTTGTCTTTTATCCAACGGATCAAGAGCTGATCATATCCGAATTTCAGAATCCAA
ATCAAGAAAAGCAAGAAGAAATCAACAGAAGAGAGTTGTTGAGCAAGTGAAAGAAGAGAA

TCTGTTGTGCGACGCATGGGCGTGGCGTAAATACGGGCAGAAACCCATCAAAGGATCTCC
ATACCCAAGGAGTTATTACAGATGCAGTAGCTCAAAAGGGTGTGTTGGCAAGAAAACAAGT
CGAAAGAAATCCTCAAAACCCGGAGAAATTCACCATAACATACTAATGAGCACAAATCA
TGAATACCAACCCGGAGAAACTCATTAGCCGGTTCGACTCGAGCAAAAACCTCCCAACC
CAAACCAACCTTAACCAAAAAATCCGAAAAAGAGTTGTTTCTCCCTACAAGTAATCC
TATGATCCCATCCGCTGATGAATCTTCTGTTGCGGTTCAAGAAATGAGCGTTGCGGAAAC
GAGTACGCACCAAGCGGCTGGAGCAATCGAGGGCCGCGCTTGAGTAACGGTTTACCATC
GGATTTGATGTCCGGGAGCGGAACTTTTCCAAGTTTTACCAGTGACTTCGATGAATATT
GAATAGCCAAGAGTTCTTCAGTGGGTATTTATGGAATTACTAGAGAGCATTAGGTGTATG
TATATATATAT

>G1756 Amino Acid Sequence (domain in AA coordinates: TBD)
MGEVAYMDEGDLEAIVRGYSGSDAFSGESSGTFSPSFCLPMETSSFYEPMETSGLDEL
GELYKFPYFSTQTILTSSVSLPEDSKPFRDDKKQRSHGCLLSNGSRADHIRISESKSKK
SKKNQQRVVEQVKEENLLSDAWARKYQKPIKGSYPYRYSYRCSSSKGCLARKQVERN
PQNPEKFTITYTNEHNHELPTRNSLAGSTRAKTSQPKPTLTKKSEKEVVSSPTSNNMIP
SADESSVAVQEMSVAETSTHQAGAIEGRRLSNGLPDLMSGSGTFPSFTGDFDELLNSQ
EFFSGYLWNY*

>G1757 (250..1224)
ATCACCAATCCTATAACACTCTCATTTCTCATCATATCATTCTTCAATCTATATAACCCAT
TCTTAATTATACTCAACACACATTATATTTTTCTGATCATATCATTCTTTCAGTCCATCT
ATATAACCAATTCTTGATTTATACTTAAACACACATTATACATCTTCTCATCATAGTT
TGTATCAATTTCTAGAGTAACTACCTAAAGGAAAAAAAATCTATTTTGGGAATCAT
ATACTAAAAATGGAAGGAAGAGATATGTTAAGTTGGGAGCAAAAGACATTGCTAAGCGAG
CTTATCAATGGATTTTGATGCGGCCAAAAAGCTTCAGGCACGACTTAGAGAAGCTCCGTGCG
CCGTGCTCATCATTTTTCATCACCGGCGACGGCTGTTGCTGAGACTAACGAGATTCTGGTG
AAGCAGATAGTTTCTTCTACGAGAGATCTCTTCTTCTGCTAAACTGGTCATCCTCACCG
AGCGTACAACCTATTCCGACGCCGTTACTGTAGTCCCGGTGGCAAATCCCGGCAGTGTT
CCAGAATCTCCGGCATCGATAAACGGAAGTCCGAGAAGTGAAGAGTTTGCCGATGGAGGA
GGTTCTAGCGAGAGTCATCATCGCCAGATTACATTTTCAATTCAAAGAAAAGAAAGATG
TTACCAAAGTGGTCAGAAAAAGTGAGAATAAGCCCAGAGAGAGGCTTAGAAGGACCTCAA
GATGATGTCTTTAGCTGGAGAAAATATGGTCAAAAAGACATTTTAGGCGCCAAATTCCCA
AGGAGTTATTACAGATGCACACATCGTAGCAACAAAACTGTTGGGCAACGAAAACAAGTC
CAGAGATCAGACGGGGATGCTACGGTTTTTCGAAGTGACGTACAGAGGAACACACACTTGT
TCGCAGGCGATCACAAGAACACCACCATTAGCCTCGCCGAGAAGCGACAAGACACCAGA
GTCAAACCAGCCATTACCCAAAAGCCAAAGGATATTCTCGAGAGTCTTAAATCCAACCTTA
ACCGTTTGAACCGATGGGCTTGATGATGGTAAAGACGTTTCTCGTTCCCTGTATACGCCG
CCGTTTTACAATTACGGAACATCAACGGCGAGTTCGGCCACGTGGAGAGTTCTCCGATC
TTCCGACGTTGTTGACTGGTTCAATCCAACGGTCGAGATTGACACAACCTTCCCCGCGTTT
TTACACGAGTCGATTTATTATTAATTAATTTGTAACAGAGAAATAGATAGTAAGT
AAGTAATGATCAGCGAGAGTTAAACATAAAAGTACTTAGAGTAATCTAACGATGCATAA
TAAGGAATGTTCAACAGGACTTGAACATGATTTCAATACTAAGAGAGATTTATCTAGCTA
CTGGTAGTAGCCGCAGACTTCTTGTGTAGCTTCACTTNCCTTTTGTGCTT

>G1757 Amino Acid Sequence (domain in AA coordinates: 158-218)
MEGRDMLSWEQKTLLELINGFDAKKLQARLREAPSPSSSFSSPATAVAETNEILVKQI
VSSYERSLLLLNWSSSPSVQLIPTVTVVPVANPGSVPEPASINGSRSEEFADGGSS
ESHHRQDYIFNSKKRKMPLKWSEKVRISPERGLEGPQDDVFSWRKYQKDILGAKFPRSY
YRCTRSTQNCWATKQVQRSDGATVFEVTRYRGTHTCSQAITRTPPLASPEKRQDTRVKP
AITQPKPDILESLSNLTVRTDGLDDGKDVFSFPDTPPFYNYGTINGEFHVESSPIFDV
VDWFNPTVEIDTTFPAFLHESIYY*

>G1782 (1..927)
ATGCAAGTGTTCAAAGGAAAGAAGATTATCTTGGGGAAACTCAATGCCTACAACAAAT
TCAAATATTCAAGGATCTGAATCTTTCAGCTTGACTAAGGATATGATAATGTCTACAACA
CAATTACCCGCGATGAAACATTCGGGTTTGCAGCTGCAAAATCAAGATTCAACCTCATCA
CAATCTACTGAAGAAGAATCAGGCGGCGGTGAAGTTGCAAGCTTTGGAGAATATAAGCGT
TATGGATGCAGCATTGTTAATAACAATCTCTCAGGTTACATCGAAAACCTTGGGAAAGCCT
ATTGAAAATTATACTAAGTCAATTACTACCTCGTCGATGGTGTCTCAAGACTCTGTGTTT
CCTGCTCTACTTCTGGTCAAAATATCTTGGTCTCTTCAATGTGCTGAAACGTCAATTTT

AATGGTTTCTTGGCTCCTGAATATGCATCAACACCAACGGCGCTGCCACATTTAGAGATG
ATGGGTTTGGTTTCTTCAAGAGTGCCATTGCCTCATCACATTCAAGAGAATGAACCAATA
TTTGTCAATGCGAAACAGTATCATGCGATTCTCCGTCGCAGGAAGCACCCTGCTAAACTC
GAAGCTCAGAACAAACTCATCAAATGCCGTAAACCGTACCTTCATGAGTCTCGCCATCTT
CATGCTTTAAAGAGAGCTAGAGGCTCCGGTGGACGTTTCTCAATACAAAGAAGCTTCAA
GAATCATCAAACCTACTGTGTTCTTCTCAAATGGCAAATGGACAAAATTTCTCTATGAGC
CCTCACGGTGGTGGAGCGGAATCGGGTCTAGTTCGATCTCACCGAGCTCCAATTCAAAC
TGTATCAACATGTTCCAAAACCCGCAGTTCAGATTCTCAGGTTATCCGTCAACACACCAT
GCCTCAGCTCTCATGTGAGGGACTTGA

>G1782 Amino Acid Sequence (domain in AA coordinates: 166-238)

MQVFQRKEDSSWGNMPTTNSNIQSESFSLTKDMIMSTTQLPAMKHSGLQLQNQDSTSS
QSTEEESGGGEVASFGEYKRYGCSIVNNLSGYIENLGKPIENYTKSITTSSMVSQDSVF
PAPTSGQISWSLQCAETSHFNGFLAPEYASTPTALPHLEMMGLVSSRVPLPHHIQENEP
FVNAKQYHAILRRRKHRAKLEAQNKLICKRKPYLHESRHLHALKRARGSGGRFLNTKKLQ
ESSNSLCSSQMANGQNFMSMPHGGGSGIGSSSISPSSNSNCINMFQNPQFRFSGYPSTHH
ASALMSGT*

>G184 (327..1937)

TGAATTCTAGCCTTTTGTAGGCGAATCATCTGGACCGGTAAGAGACTCTCTCATCGATA
ATAACCACATAATTTAATCAAACCTCTTCTCTCTCTTCTAAGATCTTTTGCTTTGCTCT
TTTCTTTTGGATCTTCTATATATGGAGAAGCACCAAAACGGTACTTACTATACGATAC
TGTACGGATCCATCAAACCTGGATTAATTATCAAAACGTACATTTTATCTTACCTGGCAA
GTTACATTCTAGGGTTTGGAGAATCCAATCAACAACAAAGAAAATAATCATCGTTACA
ATAATCAGTATCACGCACAGACTTAGATGTTCCGGTTTCCAGTGAGTCTAGGCGGTTTAC
GTGACGAAGACCGTACGATCAGATCACACCGTTGGATGACCATCGTGTGGTGGTTGATG
AGGTTGACTTCTTCTCAGAGAAGAGAGATAGGGTTTACGTTGAGAATCAACGACGACG
ACGACGAAGGCAATAAGGTTCTCATCAAATGGAGGGTTACAGAGTTGAAGAAAACGATC
GTTCCAGAGATGTCAATATCGGTCTGAATCTTCTGACCGGAATACGGGAAGCGATGAGT
CAACGGTGGATGATGGACTATCAATGGATATGGAAGATAAACGTGCAAAGATTGAGAACG
CACAACCTACAAGAAGAGCTCAAGAAGATGAAAATAGAGAATCAAAGGCTAAGAGATATGT
TGAGCCAAGCGACGACCAACTTCAATGCCTTACAAATGCAACTTGTGCGCGTCATGAGGC
AACAAGAACAACGTAACCTCTCACAAGATCATCTCCTGGAGAGCAAAGCAGAAGGAAGGA
AACGGCAGGAACCTGCAATCATGGTGCCAAGGCAGTTCATGGACCTTGGGCGCTCGTCTG
GAGCAGCAGAGCATGGAGCCGAAGTGTCTATCTGAAGAGAGGACAACGGTTCGTTCAAGTT
CTCCTCTTCGCTTCTAGAAAGTTTCCAATCCCCGAGAGAACGGAAGAGGTTGCTTGGAA
GAGAAGAAAGCTCAGAGGAATCAGAGTCTAACGCCTGGGGAAACCTAACAAAGTCCCCA
AACATAATCCATCCTCTAGCAATAGCAATGGAACAGAAACGGAATGTTATGATCAGT
CGGCCGAGAAAGCCACCATGCGGAAAGCCCGTGTCTCAGTTCGTGCCCCGATCTGAAGCTG
CCATGATAAGCGATGGATGTCAATGGAGAAAGTACGGACAAAAAATGGCTAAAGGAAACC
CGTGTCCGCGGGCTTATTATCGTTGCACAATGGCCGGTGGATGTCCAGTTTCGCAAGCAAG
TGCAGCGTTGCGCAGAAGACAGATCTATTCTCATAACCACCTACGAAGGAAACCACAACC
ATCCACTCCCACAGCCGCTACGGCCATGGCCTCAACAACCACCGCAGCTGCAAGCATGC
TCCTCTCGGGCTCAATGTCTGAGTCAAGACGGTTTAAATGAACCCAACAAACCTCCTAGCTC
GAGCTATCTTGCCTTGCTCTCAAGCATGGCTACAATCTCAGCCTCCGCACCATTTCCCAA
CCATCACATTGGACCTCACCAATTCACCAACGGTAACAACCTAATATGACCACTAATA
ACCCGTTGATGCGATTGCTCAACGGCCCGGTTTCAACCCGCGAGTTTTCGCTCAAGTGG
TTGGTCAAGCTATGTACAATAACCAACAACAGTCCAAGTTTCTGGTTTACAGTTACCGG
CTCAGCCACTGCAGATCGCGGCCACTTCCTCGGTGGCCGAGAGCGTTAGTGTGCCAGTG
CAGCAATTGCGTCCSATCAAACCTTTCGCGCGGCTCTAGCGGCAGCGATCACGTCCATTA
TGAACGGTTCCAGTCATCAAATAATAACACCAATAATAATAATGTGGCTACGAGCAACA
ATGACAGTAGGCAATAAGAGTTTTTCAATTTGATGGTCGATTTTTTTTTTTGGGG

>G184 Amino Acid Sequence (domain in AA coordinates: 295-352)

MFRFPVSLGGSRDEDRDQITPLDDHRVVVDEVDFSEKRDRVSRENINDDDDDEGNKVL
KMEGSRVEENDRSRDVNI GLNLLTANTGSDESTVDDGLSMDMEDKRAKIENAQLQEELKK
MKIENQRLRDMLSQATTNFNALQMLVAVMRQQEQNRSSQDHLLESKAEGKRQELQIMV
PRQFMDLGPSSGAAEHAESSEERTTVRSGPSPLLESSNPRENGKRLLGREESSESE
SNAWGNPNKVPKHNPSNSSNNGNRNGNVIDQSAAEATMRKARVSVRARSEAAMISDGCQW
RKYGQKMAKGNPCPRAYRCTMAGGCPVRKQVQRCADRSILITTYEGNHNHPLPPAATA

MASTTTAAASMLLSGSMSSQDGLMNPTNLLARAILPCSSSMATISASAPFPITLTLTNS
PNGNNPNMTTNNPLMQFAQRPGFNPVLPQVVGQAMYNQOQSKFSGQLQPAQPLQIAAT
SSVAESVSAASAAIASDPNFAAALAAAITSIMNGSSHQNNNTNNNNVATSNNDNRQ*

>G1845 (111..989)

AAGACATAATTTTCTCTGTTTTCTAGCTCTCTCCTCTCAAATTTCTTCCATTGCTCTCTG
TTTTGGCAAATCGTGAAGTCCACGCTCTTTAAGGCATCAGTGAAGCAAAGATGGACTTTG
ACGAGGAGCTAAATCTTTGTATTACGAAAGGTAAAAATGTTGATCATTCTTTTGGAGGAG
AAGCTTCTTCCACGTCCCCAAGATCTATGAAGAAAATGAAGAGTCTAGTCGTCCTAAAC
CCTATTTCCAATCCTCTTCTCTCTTATTCTGTTAGAGGCTTCCCTTTTTCTCTCGATC
CAACACTTCAGAATCAGCAACAACAACTCGGATCATACTTCCGGTACTTGAGCAACGAC
AAGACCCGACAATGCAAGGCCAGAAGCAAATGATCTCCTTTAGTCCTCAACAACAACAAC
AGCAGCAGCAGTATATGGCCCAGTACTGGAGTGACACATTGAATCTGAGTCCAAGAGGAA
GAATGATGATGATGATGAGCCAAGAAGCTGTTCAACCTTACATCGCAACGAAGCTGTACA
GAGGAGTGAGACAACGTCATGGGGAAAATGGGTCGCAGAGATCCGTAAGCCACGAAGCA
GGGCACGCTTTGGCTTGGTACCTTTGATACAGCTGAAGAAGCTGCCATGGCCTACGACC
GCCAAGCCTTCAAATTACGAGGCCACAGCGCAACACTGAATTTCCCGGAGCATTTTGTGA
ATAAGGAAAGCGAGCTGCATGATTCAAACCTCGTCGGATCAGAAAGAACCTGAAACGCCAC
AGCCAAGCGAGGTTAACTTGGAGAGCAAGGAACCTACCGGTGATTGATGTTGGGAGAGAGG
AAGGTATGGCTGAGGCATGGTACAATGCCATTACATCGGGATGGGGTCTGAAAGTCTCTC
TTTGGGATGATTTGGATAGTTCTCATCAGTTTTTCATCAGAAAGCTCATCTTCTTCTCTC
TCTCTTGTCTCTATGAGGCCTTTCTTTGAAAAAGTTTATAAACCCACATTGTGTGTAGG
TTATAGTTTGGGTTATGCTCATTGGCATTGGATGGAGGCAATTTTGTGATCTCCCAT
TCCACCACATATCAGTCATTATATGTGTCTACCTTTTCTCTGTATTTCTATCATTATCAT
TGTTTTTATTATGTGTCTGTATGTGTTCCCTATTGCTACATACATAGATGTCTCTTTG
TTCAAAAAAAAAAAAAAAAAAAAAA

>G1845 Amino Acid Sequence (domain in AA coordinates: 140-207)

MDFDEELNLCITKGNVDHSHFGGEASSTSPRSMKMKSPSRPKPYFQSSSSPYSLEAPPF
SLDPTLQNNQQQLGSYVPVLEQRQDPTMQGQKQMISFSPQQQQQQQYMAQYWSDTLNL
PRGRMMMMMSQEAQVQPYIATKLYRGVRQROWGKWVAEIRKPRSRARLWLGTFTAEBAAM
AYDRQAFKLRGHSATLNFPEHFVNKESELHDSNSSDQKEPETPQPSEVNLESKELPVIDV
GREEGMAEAWYNAITSGWGPESPLWDDLDSHSHQFSSSSSSPLSCPMRPF*

>G1879 (3..917)

AAATGCCCTTAGAGGCTGTCTGATACCCGCAAGATCCATTCCGATATCTCTCCAATTGCA
AAGATTTTATGTTCCACGACTTATACTCTCAAGAAGAGTTCGTAGCTCAAGATACGAAGA
ACAACATTGATAAGTTAGGGCATGAACAGAGCTTTGTGGAACAAGGTAAGGAGGACGATC
ATCAATGGCGAGACTATCATCAGTATCCTTTGTTGATCCCTTCGTGAGGAGAAGAGCTTG
GTCTTACCGCCATTGATGTGGAGAGTCACTCCTCCACAGCACCGAGGAAGAGGAGGA
GAACGAGAAACTGCAAGAACAAAGGAAGAGATCGAGAACCAGAGAACTGACTCACATCGCCG
TCGAGAGAAATCGCCGAAACAGATGAACGAGTATCTGGCTGTGCTCCGTTCTCTAATGC
CGTCTGCTGATGCTCAAAGAGGAGATCAAGCGTCGATAGTAGGAGGAGCTATAAACTACG
TGAAGGAGTTAGAGCATATTTACAATCTATGGAGCCGAAGAGAACTAGGACTCATGATC
CCAAAGGAGACAAGACTAGCACTAGCTCGTTAGTGGGTCCATTACAGATTTTTTTCAGCT
TCCCACAATATTTCTACAAAGTCATCATCAGATGTACCGAAAGCTCATCTTACCGGCGG
AGATAGAGGTTACGGTGGCAGAAAGCCATGCGAACATCAAGATAATGACGAAGAAGAAAC
CGAGGCAGCTTCTTAAGCTCATACTTCTTTACAAAGCCTAAGGCTCACTCTTCTTCATC
TCAATGTCAACACTCTCCACAACCTCATTTCTACTCCATCAGCGTCAGGGTTGAAGAAG
GAAGCCAACCTGAATACCGTGGACGACATTGCAACAGCTTTGAATCAAACCATAAGGAGGA
TTCAAGAAGAGACATAATTCAGCAAATAGATTATAATTAACCTGTTTTATTTTATTTTA
TTTTGAAATAACTGAAATCAGTTTTCTAATTTTTTTTTTTTTTCACTATTCCTCTAATCC
TCCCTATGTAAGTTGATTTTGTCTCTGTAATGAATCAATGGTCATAAAGATCTGAAC
AAAAAAATTGAATAAAAGAAAATGGTT

>G1879 Amino Acid Sequence (domain in AA coordinates: 107-176)

MPLAENVYPQDPFGYLSNCKDFMFHDLYSQEEFVAQDTKNNIDKLGHEQS FVEQKEDDH
QWRDYHQYPLLIPLSGEELGLTAIDVESHPPQHRRKRRRTRNCKNKEIENQRMTHIAV
ERNRRKQMNEYLAFLRLMPSSYAQRGDQASIVGGAINYVKELEHILQSMEPKRTRTHDP
KGDKTSTSSLVGPFDTDFSFPPQYSTKSSSDVPESSSSPAEIEVTVAESHANIKIMTKKKP
RQLLKLITSLQSLRLTLHLNVTTLHNSILYSISVRVEEGSQLNTVDDIATALNQITIRI

QEET*

>G1888 (1..729)

ATGAAGATTTGGTGTGCTGTTTGTGATAAAGAAGAAGCTTCGGTGTTTTGTGTGCGGAT
GAAGCAGCTCTTTGTAATGGTTGCGATCGCCATGTTTCATTTTCGCCAATAAACTAGCCGGG
AAACATCTCCGGTTCTCTCTCACTTCTCCTACTTTCAAAGATGCTCCTCTTGTGATATT
TGCGGGGAGAGGCGTGCATTATTATTTTGCCAAGAAGACAGAGCAATACTATGCAGAGAA
TGTGACATTCCAATACATCAAGCTAATGAGCACACTAAGAAACACAATAGATTCTCTCCTT
ACCGGCGTTAAGATCTCTGCCTCCCCGTCAGCCTACCCAAGAGCCTCCAATTCCAATCTCT
GCTGCTGCATTTGGTCGAGCCAAAACCCGACCAAAATCAGTATCGAGCGAGGTCCCAGC
TCGGCCTCCAATGAGGTATTTACGAGCTCTTCTTCGACGACCACGAGCAATTGCTATTAT
GGGATAGAAGAAAACCTACCATCACGTGAGCGATTCCGGGTCGGGATCGGGTGTACAGGT
AGTATATCCGAGTATTTGATGGAGACATTACCGGGTTGGAGAGTGGAGGATTTGCTTGAA
CACCTTCTTGTGTCTCCTATGAGGATAACATTATTACTAATAACAATAACAGTGAGTCT
TATAGGGTTTATGATGGTTCCTCACAAATTCATCATCAAGGGTTTTGGGATCACAAACCC
TTCTCTTGA

>G1888 Amino Acid Sequence (domain in aa coordinates: 5-50)

MKIWCAVCDKEEASVFCCADEAALCNGCDRHHVFANKLAGKHLRFSLTSPTFKDAPLCDI
CGERRALLFCQEDRAILCRECDIPIHQANEHTKKHNRFLLTGVKISASPSAYPRASNSNS
AAAFGRAKTRPKSVSSEVPSSASNEVFTSSSSTTSNICYGIEENYHHVSDSGSGSGCTG
SISEYLMETLPGWRVEDLLEHPSCVSYEDNIITNNNNSESYRVYDGSQFHHQGFWDHKP
FS*

>G189 (34..987)

CCCAACTCTCTCCTTGTAGAGAGAGAGATTTTATGGCGGTGGAGCTCATGACTCGGAAT
TACATCTCCGGCGTCGGAGCTGATAGCTTCGCCGTTCAAGAAGCAGCTGCTTCAGGACTC
AAAAGTATCGAAAAATTCATCGGTTTAATGTCTCGTGATAGCTTTAACTCTGATCAGCCA
TCTTCTTCTTCCGCTCCGCTCCGCTCCGCTCCGCGCCGAGATCTTGAATCAGCTCGTAAC
ACAACGGCGGACGCGGCTGTTTCAAAGTTTAAAAGAGTCATATCTCTCTTAGATCGAACT
CGAACCGGACACGCCCGGTTTAGACGTGCTCCGGTTCATGTTATTTCTCCGGTCTCTTTA
CAAGAAGAACCAAAAACGACGCCGTTTCAGTCTCCTCTTCTCCTCCGCCGCAAATGATC
CGAAAAGGTTCTGTTTCTTCATCGATGAAAACGATTGATTTCTCATCTCTCTCCTCTGTA
ACAACCGGAATCAGACAACCAAGAAGATTTCATCATCAACGTCCCTCTGAAACGGCG
CCGTTTGGCTCTCAAACCTCAAAGCCTCTCCACGACGGTCTCGTCTTCTCAAATCAACA
AAGAGAAAATGTAACCTGAGAATCTTCTCACCGGAAAATGCGCTTCCGCTTCTTCTCCTCC
GGTCTGTGTCATTGTCTCGAAGAAAAGATAAAACAGAGGAGAATAATTAGGGTCCG
GCGATAAGTGCAAAAATGTCCGATGTACCACCGGACGATTATTCATGGAGGAAATACGGA
CAAAAACCAATTAAAGGATCTCCACATCCAAGAGGATATTATAAGTGTAGTAGCGTAAGA
GGTTGTCCAGCACGTAAACATGTTGAGAGAGCAGCTGATGATTTCGTCCATGTTGATTGTT
ACTTATGAAGGAGATCATAATCATTCTCTCTCCGCCGCTGATCTCGCCGGAGCCGCCGTT
GCTGATCTTATTTTGGAAATCGTCTTGAAAAGAACAATCTTTATTTAAGGCTTTTATAAT
ATAAATTTAGATCCTTACTTAGTGAAGTACTCAAACATGAATGAAATCAATGTAATCAA
AATCAAAAAGCTTTTGTCTAAAAA

>G189 Amino Acid Sequence (domain in AA coordinates: 240-297)

MAVELMTRNYISGVGADSFVQEEAAASGLKSIENFIGLMSRDSFNSDQPSSSASASASA
AADLESARNTTADA AVSKFRVISLLDRTRTG HARFRAPVHVISP VLLQEEPKTTPFQS
PLPPPPQMIRKGSFSSSMKTIDFSSLSVTTESDNQKKIHHHQRPS ETAPFASQTQSLST
TVSSFSKSTKRKCNSENLLTGKASASSSGRCHCSKKRKIKQRRIRVPAISAKMSDVPP
DDYSWRKYGQKPIKGSPPHPRGYKYCSSVRGCPARKHVERAADDSSMLIVTYEGDHNHSL
AADLAGA AVADLILESS*

>G1939 (92..844)

AATCATTAGCTCTTCTCTTCTCTCTCTCACAGAGAGAGTAATCACAAGCCAAGTGAGA
AAAAGAAAACACTAAACCCAGATCGAAAACCATGTCTATTAAACAACAACAACAACA
CAACAATAACAACGATGGTCTTATGATCTCATCAAACGGAGCTTTAATCGAACAACAACC
ATCAGTCGTTGTGAAGAAACACCGCGGAAAGATCGACATAGCAAAGTCGATGGAAGAGG
GAGAAGAATCCGTATGCCGATTATATGTGCTGCTCGTGTTCCTCAGCTAACGAGAGAGCT
TGGTCATAAGTCAGATGGCCAAACAATGGAATGGTTACTTCGTCAAGCAGAGCCTTCTAT
TATAGCTGCAACAGGAACCTGGTACAACCTCCAGCGAGTTTCTCAACTGCTTCTGTCTCTAT
CCGTGGAGCCACCAATTCTACTTCTTTAGATCATAAACCCACTTCTTTACTTGGTGGTAC

GTCACCGTTTATACTTGGGAAACGTGTTAGAGCTGATGAGGATAGTAATAATAGTCATAA
TCATAGTTCTGTTGGTAAAGATGAGACCTTTACGACAACACCAGCTGGGTTTTGGGCTGT
TCCGGCGAGGCCGGATTTTGGACAAGTTTGGAGTTTGTCTGGAGCTCCAACAAGAGATGTT
TTTACAACAACAACATCATCATCAGCAACCATTGTTTGTTCATCAGCAACAGCAACAACA
AGCTGCAATGGGTGAAGCTTCTGCTGCTAGAGTTGGGAATTATCTTCCGGGTCATCTTAA
TTTGCTTGCTTCTTTATCCGGTGGATCTCCCGGTCGGATCGAAGAGAGGAAGATCCACG
TTAATGGTTTAAGCCCTTTTAGGTTTGAGGGCAAAATTTGGTATATATATTTATTATCTT
CTCTTCTCTATTGTTGTCTATTGTTTCTCTATGTGTGTGTTTTAGTGTGTTAGAGATTGA
TTTGGTTTCAGAATCTCTGCAAGTGATTTGAGAGTTTTCGTTAGCTTTAAGTAAGTTAAA
GACGGTTGTTTTTGATTAGGGTTAAATTAGGGTTTAAGAATCTGTTGTTTTTTTGGAGGG
AGATCGATTTCTTATCGGATCCAAGATTACTTTTAGGAAAAAAGGGAAAATTTTCAGAAAC
CACGGTGGTTTCTTTTCTCTTTTTTTTTTTTTT

>G1939 Amino Acid Sequence (domain in AA coordinates: 40-102)
MSINNNNNNNNNNDGLMISSNGALIEQQPSVVVKPPAKDRHSKVDGRGRIRMPIICA
ARVFQLTRELGHKSDGQTIEWLLRQAEPSSIIAATGTGTTTASFSSTASVSIRGATNSTSLD
HKPTSLGGTSPFILGKRVRADEDSNNSHNHSSVGKDEFTTTTTPAGFWAVPARPDFGQVW
SFAGAPQEMFLQQQHHHQPLFVHQQQQQQAAMGEASAARVGNYPGLHNLASLSGGSP
GSDRREEDPR*

>G194 (192..1205)

TCTTTCTCTCTCTCTATCTCTCTCTTTGAACCCTAAAACTCTTTCTTTACAAGGATT
GATCTTTTTGTATTTTGTATTTTGACATTGCTTTGTGTTTCGATCTCTGTTTTGATGCGA
TTTCTCTGTTTTTAAAGCCATTGATAGATTGTTTCCGGTAAAGCTCAGCGAGAGAAGAA
GAAGAACAACAATGGAGTTTACAGATTTCTCAAAGACGAGTTTTTACTACCCGTCGTCAC
AAAGCGTTTGGGATTTTCGGAGATTTAGCGGCGGCGAGAGGCATTCTTAGGGTTCATGG
AGTTATTAAAGTTCTCAGCAGCATCAAGACTTTGCTACTGTTTCTCTCATTCTCTCTTC
TCCAAACGCTCTCAACCGCAAACGCAAACGCAACCATCGGCGAAGCTGTCTTCAAGTATCA
TTCAAGCTCCACCGTCAGAGCAATTAGTGACGTCAAAGGTGGAGTCTTGTGTTTCGGATC
ATTTGTTGATAAACCCACCGGCGACTCCTAACTCGTCATCGATTTCGTCTGCTTCAAGCG
AGGCTCTAAATGAAGAGAAACCGAAAAACAGAAGACAATGAAGAAGAAGGAGGTGAAGATC
AACAAGAGAAGAGTCATACTAAGAAACAGTTGAAAGCAAAGAAGAATAATCAGAAGAGAC
AGAGAGAGGCAAGAGTCGCATTTCATGACAAAGAGTGAAGTTGATCATCTCGAAGATGGTT
ATCGCTGGCGAAAAATATGGTCAAAAAGCTGTCAAACAGTCCTTTTCCAGGAGTTACT
ACCGTTGCACAACGGCTTCATGTAACGTGAAGAAGAGAGTGGAGAGATCATTTCAGAGATC
CAAGCACTGTGGTTACAACCTACGAAGGTCAACACACTCACATTAGTCCACTCACGTCTC
GTCCTATTTCCACTGGAGGTTTCTTCGGATCGTCAGGAGCTGCTTCGAGTCTCGGTAATG
GTTGCTTTGGGTTTCTATTGATGGCTCCACGTTAATCTCTCTCAGTCCAACAGCTTG
TCCAATACCATCACCAACAGCAGCAACAAGAATCATGTCTTGTGTTTGGAGGAGTCAACG
AGTACCTTAATAGCCACGCTAATGAGTATGGTGATGATAATCGTGTGAAGAAGAGTCGAG
TTTTGGTTAAAGATAATGGACTTCTGCAAGATGTTGTTCCGTCTCATATGTTGAAGGAAG
AGTAGTAGTATATATATAGTCTTATAGTTTTAATCTAGTTTTTTTTTGTATAATTGTCTA
AAAGAAACGGATCTTTTGTTCGATGAGAAGATGTTTCTTATGGTTCTGAAATCGTAA
GGTAATGATGATTGTACCAAGCCGAGAAAGTACTTGTGATTTTACCATTGAATCACTAT
AAATGTAATTTTTTATTACTGTGAAAAAAAAAAAAAAAAA

>G194 Amino Acid Sequence (domain in AA coordinates: 174-230)
MEFTDFSKTSFYYPSSQSVWDFGLAAAERHSLGFMELLSSQQHQDFATVSPHSFLLQTS
QPQTQTQPSAKLSSSIIQAPPSEQLVTSKVESLCSHLLINPPATPNSSSISSASSEALN
EEKPKTEDNEEEGGEDQEQESHKKQLKAKKNNQKRQREARVAFMTKSEVDHLEDGYRWR
KYGQKAVKNSFPFRSYRCTTASCNVKKRVERSFRDPSTVVTTYEQHTHISPLTSRPIS
TGGFFGSSGAASSLNGCFGFPIDGSTLISPQFQQLVQYHHQQQQELMSCFGGVNEYLN
SHANEYGDDNRVKSRVLVKDNGLLQDVVPSHMLKEE*

>G1943 (137..1858)

ACATTTGTTTCTAATCTCAGACATAAATAATTTTTGTTCGGACTTCAAAACCAACGATG
ATTATATCATTCACATTCTTTCTTCTACTTCTTCTCTCTCTGATCTCATTTCCTT
AGAAAATCCATCTATCATGGGTGAAGATGATATAGTGGAGCTCTTATGGAAGAGTGGCCA
AGTCGTTAGAACCAGTCAAACAGAGACCCTCCTCCAATACACCACCATCTCTCTCTCC
ACCACCCATTCTTCGTGGTAGCGGAAGCGGCAACGGAGAAGAAAATGCCCCGCTTCCACT
TCCACAGCCTTCACCTCCCTCCATCATCAGAATCTTTTCATTCTGGAAGACGAAATGTC

TTCTTGGCTTACCATTCTACCCCGGCGTTACGTCCACCCCGGCTTCTTCTGTCTCCCT
GCCACCACCACCCAATGCTCCGCGTGAAGATGATATAGTGGAGCTTTTATGGCAAAGCGG
CCAAGTAGTTGGAACCAACCAACACATAGACAATCCTACGATCCTCCTCCCATTTCTCCG
CGGCAGCGGAAGTGGCAGAGGAGAAGAAAATGCTCCCTTTTACAACCTCCGCTCACCT
GCATCAGCAAAATCTCTTCAATCAAGAAGGCGAAATGTATTCTGTGGCTACACCATTCTTA
CCGCCAAAATATTTCTGTCTCAGAACTTCTCAACTCCACTCCGGCTACTCACCCGCAAAG
TTCCATCTCTCTGGCACCACGTGAGACTATCGCCACGAGAAGGGCGGAAAACCTTTATGAA
CTTCTCGTGGCTAAGAGGGAACATATTTACCGGCGGTAGAGTTGATGAAGCTGGACCGTC
GTTTTCTGGTGGTAAGAGAATCGATGCAGGTAGGCTCGAACACGACCCCCCTTCTCTTC
TGCCACTGAATCATGTGTAATACCAGCTACAGAGGGCACCAGGAGTCGAGTGTGGGAAC
TTTGGCAGCTCATGATCTTGGTCGGAAGGGAAGGCGGTGGCGGTTGAGGCGGCCGGAAC
ACCATCTTCAGGAGTGTGCAAGGCCGAAACAGAGCCGTTTCAGATACAACCAGCAACGGA
GTGCAAGCTAAAAGCGAGAGAAGAAACCCATGGAACGAAGAAGCTCGTGGTTCAACGTC
TAGAAAGAGATCACGAACTGCAGAAATGCATAACCTCGCCGAAAGGAGAAGGAGAGAAAA
GATCAACGAGAAGATGAAGACTCTGCAACAACCTCATTCTCGCTGCAACAAGGTTGAATC
TGATTCTGTTTTCTACTCTGATCAGTCTACTAAAGTTTCAACGCTGGATGATGCTATCGAG
TACGTCAAATCGTTACAGAGCCAAATACAAGTATGCTCTTCAAAACAGAATGTGTTTAA
ACCAATGGTTCAACATGGAAGAGTTCATATGTATCTAGTTTTGTTGAGATGATGTCGAC
GGGACAGGGTATGATGTCGCAATGATGAATGCCGGAATACGCAACAGTTTCATGCCCA
TATGGCCATGGATATGAACCGACCTCCTCCATTCATACCTTTCCCGGCACATCTTTTCC
TATGCGCGCTCAAATGCGAGGTGTAGGTCCATCATATCCAGCACCGCGCTACCTTTTCC
CAACATTCAGACCTTTGACCCATCCAGAGTCCGTTTACCAAGCCCGCAGCCTAACCCGGT
GTCGAACCAGCCTCAGTTTCCGGCTTACATGAATCCCTATAGCCAGTTTGCTGGTCCCA
CCAGTTGCAACAACCTCCTCCTCCTCATTTCAGGGTCAAACAACATCACAACAGTTTC
CGGGCAGGCAAGTAGTAGCAAGGAACCTGAGGATCAGGAGAACCACCAACAGCTTAGTT
AAAGTGTGGAGCTGAAACGGATCAGTTCTTCAAGCAAATTACAACCTTTGAAGATAAACCA
GAGTTGTAACATGTAGATTTTGTCTGTTAAGTTTAAATGTAAGTACTTTTAGTTAATGGG
AAAGATACTGACAGGTTGCAAGGTGGTCAGTATTTGTGCATCACGCTTAAGATTCTCTGA
TGTGGCCAGTATCTCCCTTTTCTAGCATGTGAGGTCCCTACTCTCTGGTTCTACGGAGAC
CAAATGTTTCGACTGATTAACACACAATGACTTACCAAAAGTACACGCGGCCCATCCTCG
TCTTTATGTTCCAAGTGCAGCTGTTTGTATTGTAAGCATTCTTTCTATAATAATAAA
ACAGCTCTATCTTCGTTAAAAAAA

>G1943 Amino Acid Sequence (domain in AA coordinates: 335-406)

MGEDDIVELLWKSGQVVRTSQTRPSSNTPPSLPPPPILRGSGSGNGEENAPLPLPQSP
PLHHQNLFILEDEMSSWLHSHHPGVTSTPASSVSLPPPPNAPREDDIVELLWQSGQVVG
NQTHRQSYDPPPILRGSGSGRGEENAPLSQPPPHLHQNLFIQEGEMYSWLHHSYRQNYF
CSELLNSTPATHPQSSISLAPRQTIATRRANFMNFSWLRGNIFTGGRVDEAGPSFSVVR
ESMQVGSNTTPSSSATESCVIPATEGTASRVSGTLAAHDLGRKGKAVAVEAAGTPSSGV
CKAETEPVQIQPATESKLKAREETHGTEEARGSTSRKRSRTAEMHNLAEERRRREKINEKM
KTLQQLIPRCNKVESDSVSTLISLLKFQRWMLSSSTSNRYRAKYKYALQNRMCFKPMVQH
GKSSYVSSFVEMMSTGQMMSPMMNAGNTQQFMPHMMAMDMNRPPPIFPFGTSFPMQAQ
AGVGPSYPAPRYPPFNIQTFDPSRVRLPSPQPNPVSNQPPPAYMNPYSQFAGPHQLQPP
PPPPFQGQTTSQLSSGQASSSKEPEDQENQPTA*

>G21 (79..966)

TGTGGAGGAATATTAATACAGCCCACTTCACATCTATTTTTGTGCAACCATCTCTCTAAA
GCTTCTTCTCTCATAACAATGGCAAGACAAATCAACATAGAGAGTAGTGTCTCAAGTT
ACCTTTATCTCCTCCGCCATCCCCGCCGTATCTTCTCCTCCTCCATCACCGCTTCCGCC
TCATTGTCTCTTCACTACTACATCTTCTCTTCTCGTCATCAACAAATCTAACTTC
ATTGAGGAAGACAACCTATAAAGAAAAGCATCTCGAAGATCATTGTCTATCGTTAGTCTCC
GTTGAAGACGATGATGATCAAAACGGTGGAGGTGGGAAACGGCGAAAGACCAACGGTGGA
GATAACATCCGACGTATAGAGGAGTGAGGATGAGGAGTTGGGGAAAATGGGTGTCGGAG
ATTAGAGAGCCGAGAAAGAAATCAAGAATCTGGCTCGGGACTTATCCAACGGCTGAGATG
GCAGCTCGAGCTCATGACGTAGCGGCTTTAGCCATTAAAGGTACAACGGCTTACCTCAAT
TTTCCCAAGTTAGCCGGCGAGCTTCTCGTCCAGTCACAAATCTCCTAAAGACATTCAA
GCCGCCGCCTCTTTAGCGGGCGGTTAACTGGCAAGATTCCGTTCAACGATGTGAGTAATTCT
GAAGTGGCTGAAATAGTTGAAGCCGAGCCGAGTCGAGCCGTGGTGGCTCAGTTGTTTTCT
TCGGACACAAGCACGACGACGAGCTCAGAGTCAAGAGTATTCGGAAGCTTCGTGTGCT

TCGACTTCGGCGTGTTACGGACAAAGACAGTGAGGAAGAGAAGCTGTTTGATTGCGCGGAT
TTGTTTACCAGATGAGAATGAGATGATGATACGAAACGATGCGTTTTGCTACTACTCGTCC
ACGTGGCAGCTTTGTGGAGCCGATGCTGGGTTTCGGCTTGAAGAGCCGTTTTTCTATCT
GAATGACTAAAGTACCCCTCTCGAGAGAGCTCTCACTAACACT

>G21 Amino Acid Sequence (domain in AA coordinates: 97-164)
MARQINIESSVSQVTFISSAIPAVSSSSSITASASLSSSPTSSSSSSSTNSNFIEEDNS
KRKASRRSLSSLVSVEDDDDDQNGGGKRRKTNGGDKHPTYRGVVRMRSWGKWWSEIREPRK
KSRIWLGTYPTAEMAARAHDDVAALAIKGTAYLNFPLAGELPRPVTNSPKDIQAAASLA
AVNWQDSVNDVSNSEVAEIVEAEPSRAVVAQLFSSDTSTTTTTQSQEYSEASCASTSACT
DKDSEEEKLFDLPDLFTDENEMMIRNDAFCYSSSTWQLCGADAGFRLEEPFFLSE*

>G2132 (42..1031)

ATTCTGTACTTAGTACCGGAGTTTAGTCGGAGAGAGACAATGATCAGTTTCAGAGAAG
AGAACATCGATCTCAACTTGATTAAACAATTAGTGTATCTGTAATGATCCAGACGCCA
CCGATTCTCTAGCGACGATGAATCTATCTCCGGCAATAATCTCGCCGTCAGATCAAAC
CAAAACCACCGAAACGTTACGTCTCAAAGATCTGTGTCCGACGCTGATCAAAGGTATG
AGAACGTTTTCGAATTCTACAGGGAAATAAGCAGCCGGAAACCGGAAACGTCGTCGGGTT
TCAAAGGCGTACGACCGAGGCCGTGGGGGAAATTTGCGCGGAGATAAGAAAATCCGTTTG
AGAAGAAGAGAAAAGTGGCTTGAACGTTTCTACTGAAGAAGAAGCAGCAGAAGCTTACC
AAAAGAGTAAAAGAGAGTTTGATGAACGATTGGGTTTAGTTAAACAGGAAAAAGACCTAG
TAGATTTGACCAAGCCGTGCGGTGTACGTAAACCAGAAGAGAAGGAAGTTACTGAGAAGT
CGAATTGCAAAAAGGTAAATAAGAGAATTGTTACTGATCAGAAGCCATTGTTGTGGTT
ATAACGCTGATCATGAAGAAGAGGGAGTGATTAGTAAATGTTGGAAGATCCGTTGATGA
CATCGTCAATTGCTGATATTTTTGGTGATTGCGCTGTTGAAGCAAATGATATTTGGGTGG
ATTACAATTCAGTGAATTTATTTCCATTGTAGATGATTCAAGTTTGATTTTGTGGAGA
ATGATAGAGTAGGAAAGGAGAGAAAACATTTGGATTAAAGATTGGGGATCACACTAAAGTTA
ATCAACATGCCAAAATCGTATCGACCAATGGGGACTTATTCGTCGATGATTTACTTGATT
TTGATCCGTTGATAGATGATTTTAAGTTAGAAGATTTTCTATGGATGATCTTGGATTAT
TAGGAGATCCAGAGGATGATGATTTTAGTTGGTTTAATGGTACTACTGATTGGATCGATA
AGTTTTTATGAATACTTTCTTGACACGCCAACGGTATTAGTAC

>G2132 Amino Acid Sequence (domain in AA coordinates: TBD)
MISFRENIIDLNLIKTISVICNDPPATDSSSDDESISGNPNRRQIKPKPPKRYVSKICVP
TLIKRYENVSNSTGNKAAGNRKTSSGFKGVRRRPWGKFAAEIRNPFKKRWLGTFTPTEE
EAAEAYQKSKREFDERLGLVKQEKDLVDLTKPCGVVRKPEEKEVTEKSNCKKVNKRIVTDQ
KPFPGCGYNADHEEERGVISKMLEDPLMTSSIAIDIFGDSAVEANDIWVDYNSVEFISIVDDF
KFDFVENDRVGKEKTFGFKIGDHTKVNQHAKIVSTNGDLFVDDLLDFDPLIDDFKLEDFP
MDDLGLLGDPEDDDFSWFNGTTDWIDKFL*

>G2145 (1..777)

ATGGACGTTTTTGTGTGATGGTGAATTGGAGTCTCTCTTGGGGATGTTCAACTTTGATCAA
TGTTTCATCATCTAAAGAGGAGAGACCGCGAGACGAGTTGCTTGGCCTCTCTAGCCTTTAC
AATGGTCATCTTCATCAACATCAACACCATAACAATGTCCTTATCTTCTGATCATCATGCT
TTCTTGCTCCCTGATATGTTCCCATTTGGTGCAATGCCGGGAGGAAATCTCCGGCCATG
CTTGATTCTTGGGATCAAAGTCATCACCTCCAAGAAACGTCCTCTCTTAAGAGGAAACTA
CTTGACGTGGAGAATCTATGCAAAACTAACTCTAACTGTGACGTCACAAGACAAGAGCTT
GCGAAATCCAAGAAAAAACAGAGGGTAAGCTCGGAAAGCAATACAGTTGACGAGAGCAAC
ACTAATTGGGTAGATGGTCAGAGTTTAAGCAACAGTTTCAGATGATGAGAAAGCTTCGGTC
ACAAGTGTTAAAGGCAAAACTAGAGCCACCAAGGGACAGCCACTGATCCTCAAAGCCTT
TATGCTCGGAAACGAAGAGAGAAGATTAACGAAAGGCTCAAGACACTACAAAACCTTGTC
CCAAACGGGACAAAAGTCGATATAAGCACGATGCTTGAAGAAGCGGTCCATTACGTGAAG
TTCTTGACGCTTCAGATTAAGTTGTTGAGCTCGGATGATCTATGGATGTACGCACCATTG
GCTTACAACGGCCTGGACATGGGGTTCCATCACAACCTTTGTCTCGGCTTATGTGA

>G2145 Amino Acid Sequence (domain in AA coordinates: 166-243)
MDVFDGELESLLGMFNFQDCSSSKEERPRDELGLLSSLYNGHLHQHNNVLSDDHHA
FLLPDMFPFGAMPGGNLPAMLDSDQSHHLQETSSLKRKLLDVENLCKTNSNCDVTRQEL
AKSKKKQRVSSESNTVDESNTNWVDGQSLSNSSDDEKASVTSVKGKTRATKGTATDPQSL
YARKRREKINERLKLQNLVPNGTKVDISTMLEEAVHYVKFLQLQIKLLSSDDLWMYAPL
AYNGLDMGFHNNLLSRLM*

>G23 (22..732)

TATCAAACGAGAGTACAAAAGATGACGTCACTCAACAGCTCTGCATCACCAACATCATCG
 TCATCAGACCAATCTGATGCACTACTACAACAAGCACCCACTTGTCTGAAGAAGAAGCT
 CCACCCAGAAAACAACAACAAGAAAGAGAAGGAGAGATTCTTCTTGCTTCTTCATCT
 TCTTCAATGCAACATCCTGTTTACAGAGGTGTGCGGATGAGAAGTTGGGGCAAATGGGTC
 TCCGAGATCCGACAACCTCGTAAGAAAACCTCGTATTTGGCTCGGCACTTTTGTACCCGCT
 GATATGGCTGCTCGTGCTCACGACGTGCTGCTCTCACCATCAAAGGCTCCTCCGCCGTC
 TTAAATTTCCCTGAGCTTGCTTCTCTCTTCCCTCGTCCGGCGTCATCATCGCCGCATGAT
 ATCCAGACAGCCGCCGAGAAGCCGCCGCATGGTGGTGAAGAAAACTGTTAGAGAAG
 GATGAGGCTCCGGAGGCCCCACCTTCGTGGAATCTTCTTACGTGGCGGCGAGTCAGAG
 GATGAGGAGAGTTGGAGAAAATTGTGGAGCTGCCTAACATTGAAGAAGGAAGTTATGAC
 GAGAGTGTGACATCACGTGCTGATCTGGCTTATTCTGAGCCGTTGATTGTTGGGTGTAT
 CCTCCGGTTATGGATTTTTATGAAGAAATATCGGAGTTTAAATTCGTGGAATTGTGGAGC
 TTTAATCACTAATTAAGTTAGGAAAGTGCATTATATTGCAATATTGCATCATAGATAACA
 TTTGTATTTCTTTCTTTTGTACGGATACGTAGCATATGCTACTATACTAGGGCTAGTG
 TACCAAATATTGTAAAATATACTTATTAATATTTATGTAAATGTGTAATATATATAACAT
 ACAATTATTGTAAGTTTGGAAATTGGAAGTATCGTTACGCAATGTTCTTGTAACAAAAA
 AAAAAA

>G23 Amino Acid Sequence (domain in AA coordinates: 61-117)
 MTSLSNSSASPTSSSSDQSDATTTTTSTHLSSEEEAPPRNNNTRKRRRSDSSSSSSMQHPV
 YRGVRRMSWGKVVSEIRQPRKKTRIWLGTFTADMAARAHDAALTIKSSAVLNFPELA
 SLFPRPASSSPHDIQTAAABAAAMVVEEKLLEKDEAPEAPPSSESSYVAAESEDEERLEK
 IVELPNIEEGSYDESVTSRADLAYSEPFDCWVYPPVMDFYEEISEFNFVELWSFNH*

>G2313 (104..724)

CGTCGACACAATCGCTCTTCCGTAACATATTCCACAAAACGATCTTCTGTTTCTTGAAT
 TTTTAGCCATCTCTTTTTTTTTTCTCATTCTCTCGGATACTATGGCTTCGAGTCCACG
 CTGGACGGAGGACGACAACAGGCGTTTTAAGTCAGCTCTGTGCGCAATCCCTCCGGATAA
 CAAGCGTTTGGTGAATGTGCGCCAGCATCTGCCGAAACCTTTGGAGGAGGTGAAGTACTA
 CTACGAAAAGTTGGTCAACGATGTTTATCTGCCGAAACCTTTAGAGAATGTCACCCAGCA
 TCTGCAGAAACCTATGGAATGGAGGAGATGAAGTACATGTACGAAAAGATGGCCAACGA
 TGTTAATCAGATGCCCGAGTACGTACCACTGGCGGAATCGAGTCAGTCCAAACGCAGGAA
 GAAGGATACGCCAAATCCTTGGACAGAAGAGGAACACAGATTGTTTCTGCAAGGATTGAA
 AAAGTATGGGGAAGGAGCTTCGACGTTGACATCAACGAATTTTGTGAAGACAAAGACTCC
 ACGGCAAGTGTCAAGCCATGCACAGTATTACAAAAGGCAAAAATCGGACAATAAGAAGGA
 GAAACGCCGGAGTATTTTGGACATAACTTTGGAGTCTACCGAGGGCAATCCAGATTCTGG
 AAATCAGAACCCTCCGGATGATGATGATCCGTCCCAAGGTCAAGGCACTTGTCTTGGAGT
 TTAGATGTTGGAAGATAGAAGATGGTGTGAAAGC

>G2313 Amino Acid Sequence (domain in AA coordinates: TBD)
 MASSPRWTEDDNRRFKSALSQFPDPNKRLLVNVAQHLPKPLEEVKYYYEKLVDVYLPKPL
 ENVTOHLQKPMEMEEMKMYEKMANDVNQMPEYVPLAESSQSKRRKDTNPNWTEEEHRL
 FLQGLKKYGEASTLTSTNFVKTKTPROVSSHAQYYKQKSDNKKERRSIFDITLESTE
 GNPDSGNQNPDDDDPSQGGTCLGV*

>G2344 (1..573)

ATGACTTCTTCAATCCATGAGCTTTCTGATAACATTGGAAGTCATGAGAAGCAAGAACAG
 AGAGATTCTCATTTCCAACCACCAATCCCTTCTGCAAGAAATTATGAATCAATTGTTACA
 AGTTTAGTCTACTCAGACCCGGGACTACAAATCCATGGCACCTGGACAATATCCATAT
 CCAGATCCTTACTACAGAAGCATATTTGCACCGCCTCCACAACCGTATACCGGGGTACAT
 CTACAGTTGATGGGAGTGCAGCAACAAGGCGTTCTTTTACCATCTGATGCAGTCGAGGAA
 CCTGTTTTTGTAAEGCAAAGCAATACCACGGTATACTAAGGCGCAGACAATCAAGAGCA
 AGACTTGAGTCTCAGAATAAAGTCATCAAGTCACGTAAGCCGTATTTGCATGAATCTCGG
 CATTGTCATGCGATAAGACGACCAAGAGGATGTGGCGGGCGGTTTCTAAATGCCAAGAAG
 GAGGATGAGCATCACGAAGACAGTAGTCATGAAGAAAAATCCAACCTTAGCGCTGGTAAA
 TCCGCCATGGCTGCTTCTAGTGGTACATCTTGA

>G2344 Amino Acid Sequence (domain in AA coordinates: TBD)
 MTSSIHESLSDNIGSHEKQEQRDSEHFQPPIPSARNYESIVTSLVYSDPGTTNSMAPGQYPY
 PDPYYSIFAPPPQPYTGVHLQLMGVQQGVPLPSDAVEEPVFNQKQYHGILRRRQSRA
 RLESQNKVIKSRKPYLHESRHLHAIRRPGRGGRFLNAKKEDEHHEDSSHEEKSNLISAGK
 SAMAASSGTS*

TCCTCACTCTCTCTCTTTTCTCTAACCATAAAATCTCTTTGATCTCTTTCTCTGTGTTT
TGATAATGGAAAATGTTGGTGTGTGGGATGCCGTTTACGATTTAGGGCAAACAGGGTTT
ACCCACTCTTGTCTGATTTCCACGATTTATCCGCGGAGAGGTATCCGGTAGGGTTCATGG
ATTTACTGGGTGTTTCATCGTCATACACCCACCCCATGACGCCGTGTGATGCATTTCCGACCA
CACCTAACTCGTCTCGAGCGAAGCTGTGAATGGAGATGACGAAGAAGAAGATGGAG
AAGAACAGCAGCATAAGACAAAGAAGCGGTTTAAATTACGATAAAATGAGTAGAAGCAGA
CGAAGAAGAAGGTGCCAAAAGTGTCAATTCATCACGAGGAGTGAGGTTCTTCATCTAGATG
ATGGTTATAAGTGGAGAAAATACGGTCAAAAACCTGTCAAAGACAGCCCTTTTCCAAGAA
ATTATTACCGTTGCACAACAATTGGTGTGACGTGAAGAAGAGAGTAGAGAGATCATTC

GTGATCCAAGCAGTGTAATCACCACCTTACGAAGGTCAACATACTCATCCTCGTCCACTAC
TCATCATGCCCCAAGAAGGCAGCTCTCCATCCAATGGCTCAGCTTCTAGGGCCACATTG
GCCTCCCTACACTCCCTCCTCAGCTTTTAGATTACAACAACCAACAACAAGCGCCGT
CTTCTTTTGAACCGAGTACATTAAACAGGCAAGAAAAAGGAATTAATCATGATGATGATG
ACGATCATGTTGTGAAGAAGAGTCAAGTCCGGATCTGCTGGATGGAGCTGGTTTAGTCA
AAGATCATGGCCTTCTTCAGGATGTTGTTCCCTCTCATATCATTAAAGGAAGAGTATTAGT
TAATCGCATAATTATGTAGCTAGCTAGCTAG

>G2517 Amino Acid Sequence (domain in AA coordinates: TBD)
MENVGVMGPFYDLGQTRVYPLLSDFHDLAERYPVGFMDLLGVHRHTPTHTPLMHFPPTP
NSSSSSEAVNGDDEEEEDGEEQHKTKKRFKFTKMSRKQTKKKVPKVSFTRSEVLHLLDDG
YKWRKYGQKPKVDSPPFRNYRCTTTWCDVKKRVERSFSDPSSVITTYEQHHTHPRPLLI
MPKEGSSPSNGSASRAHIGLPTLPQLLDYNNQQQAPSSFGTEYINRQEKGINHDDDDDD
HVVKKSRTDRLDGLVKGHLLQDVVPSHIIKEY*

>G2521 (103..768)
ATTCTCCACAATTTTCATAACTTTCTTCCGCTCAACTTCAGATAAATTCGGATTCTGTAGC
TCTTTCAATACGACTGCGGAGATCAGAGCCAATTATTTGGTTATGGCGTCTCTGATCTCA
GATATTGAACCGCCGACGAGTACTACTTCAGATCTCGTTCGGAGAAAGAAGAGATCCTCT
GCTTCATCCGCCGATCGTCTCGTTCAAGCGCATCTTCCGCTCTCCGGTGAGATTCACGCG
CGATGGCGATCGGAGAAGCAACAACGGATCTACTCAGCCAAACTGTTCCAAGCGCTCCAA
CAAGTCCGCTCAACTCTTCCGCTCAACATCATCATCTCCAACGGCTCAGAAACGAGGA
AAGGCGTCCGTGAAGCGCGCATCGAGCTCTTGCCGTTTCCGCTCGGGGAAGAACAATC
TGGAGCAGAGCGATCTTAGCTAATCGGATCAAACTGAAATTCGTAAACAGAGACGTCCT
CGAGCTACGATGGCGATTCGGCCATGACTACGGTGGTTAGTAGCAGCAGCAACAGATCG
AGAAAACGAGAGTGTGGTGTGAGATTGAATAAGAAGAGTATACCGGATGTTAACCGG
AAAGTACGTGTTCTAGGCCGGTTAGTTCCCGGTTGCGGTAAACAATCCGTACCGGTGATT
CTAGAAGAAGCAACTGATTATATTACAGGCTCTGGAGATGCAAGTGAGAGCCATGAATCT
TTAGTTACGCTTCTCTCTCTACGGCTCAGCTCCTCCACCGATTGATGAGGTTAAAAT
CGTCTTTTAAATCTACCATCTCTCGATCTTTCACAGCTTATGTGTATATAGAAGATTCTG
GTTTGATTATAATCTGTAATACTCTTCCCAACCGCTGATTCTTCTCTGCTACAAGTAAA
AGTAAATTTGAACCGAGTCTTCCCATTTTACGATCCTCAAGTCTAAATTAAGTATATG
ATTGATTAATAAAGTCTTTACCATTAGGGTTC

>G2521 Amino Acid Sequence (domain in AA coordinates: 145-213)
MASLISDIEPPTSTSDLVRRKKRSSASSASSRSSASSVSGEIHARWRSEKQQRITYSAK
LFQALQQVRLNSSASTSSSPTAQKRGKAVREAADRALAVSARGRTLWSRAILANRIKLKF
RKQRRPRATMAIPAMTTVVSSSSNRSRKRRVSVLRLNKKSIIPDVNRKVRVLGRLVPGCGK
QSVPVILEBATDYIQALEMQVRAMNSLVQLLSSYGSAPPPI*

>G258 (60..983)
AGTGACCACCCTGCTGGTTAATCAACACCAAGAGACCTTGTAATATATAAGTTAGGAAGA
TGAGAGAGAAGTGGGAAATGAAAAGAGATGAAATGGGACATCGATGTTGTGGAAAACACA
AAGTGAAGAGAGGTCTTTGGTCTCCAGAGGAAGACGAGAAGCTTCTTCGTTATATCACCA
CTCATGGTCATCCTAGTTGGAGTTCGTTCCAAAGCTTGCCGGGTTGCAGAGATGTGGGA
AGAGTTGCAGATTAAGGTGGATAAACTATCTAAGGCCTGATCTGAGGAGAGGTTTCGTTTA
ATGAGGAAGAAGAGCAGATTATCATCGACGTACATCGTATTCTTGGTAACAAATGGGCTC
AGATTGCTAAGCACTTACCTGGACGCACTGATAATGAAGTCAAGAACTTTGGAACTCAT
GCATTAAGAAGAAACTTCTTCTCAAGGCTTAGATCCTTCTACACATAATCTTATGCCTT
CACACAAAAGATCTTCTTCTTCAAACAATAATAATCCCCAAGCCAAACAAAACGACGT
CCATCATGAAGAACCCTACTGATCTTGATCAATCAACCACTGCTTTTCAATCACAAACA
TCAATCCACCCACTTCCACTAAACCAAACTTAAATCTCCTAACCAGACTACAATCC
CATCTCAAACCGTGATCCCTATCAATGATAACATGTCAAGTACTCAAACCATGATCCCTA
TCAATGATCCCATGTCAAGTCTTTTAGATGATGAGAATATGATTCCTCACTGGTCAGATG
TTGATGGAATGGCGATCCACGAAGCTCCGATGTTGCCTAGTGATAAGGCAGTAGTGGGAG
TGGATGATGATGATCTCAACATGGACATTTTGTTTAACTACTCTTCTTCTGCTTTTG
ATCTGATTTTGTCTTCAATTTCTCCTCTGCAATGTCTATCGATTCTCAATCCCATGGATG
ATCTTGGCAGCTGGACCTTTAGCTTTTACTCTACAGC

>G258 Amino Acid Sequence (domain in AA coordinates: 24-124)
MREKWEKMRDEMGRCCGKHVKRGLWSPEEDEKLLRYITTHGHPSSVVKLAGLQRCG
KSCRLRWINYLRPDLRRGSFNBBEEQIIIDVHRILGNKWAQIAKHLPGRTDNEVKNFWNS

CIKKKLLSQGLDPSTHNLMPSHKRSSSSNNNNI PKPNKTTSIMKNPTDLDQSTTAFSITN
INPPTSTKPNKLKSPNQTTIPSQTVIPINDNMSSTQTMIPINDPMSSLLDDENMIPHWS
VDGMAIHEAPMLPSDKAVVGVDLNDILFNTPPSSSAFDPDFASIFSSAMSIDFNPM
DLGSWTF*

>G280 (108..722)

AAGTTAATATGAGAATAATGAGAAAACCACTTTCCCAAATGCTTTTTTAAATCCCTCCT
CACACAGATTCCCTTCCTTCATCACCTCACACTCTCTACGCTTGACATGGCCTTCGATC
TCCACCATGGCTCAGCTTCAGATACGCATTATCAGAACTCCGTCGTTTTCTCTCCAC
CTTATCCTCAGATGATAATGGAAGCGATTGAGTCCTTGAACGATAAGAACGGCTGCAACA
AAACGACGATTGCTAAGCACATCGAGTCGACTCAACAACTCTACCGCCGTCACACATGA
CGCTGCTCAGCTACCATCTCAACCAGATGAAGAAAACCGTCAGCTAATCATGGTGAAGA
ACAATTATATGAAAACAGATCCAGATGCTCCTCCTAAGCGTGGTCGTGGCCGCTCCTCCGA
AGCAGAAGACTCAGGCCGAATCTGACGCCGCTGCTGCTGCTGTTGTTGCTGCCACCGTCG
TCTCTACAGATCCGCTTAGATCTCGTGGCCGCTCACCAGCCGAAAGATCCATCGGAGC
CTCCCAGGAGAAGGTCATTACCGGATCTGGAAGGCCACGAGGACGACCACCGAAGAGAC
CGAGAACAGATTTCGGAGACGGTTGCTGCGCCGGAACCGGCAGCTCAGGCGACGGTGAGC
GTAGGGGACGTGGGAGACCTCCGAAGGTGAAGCCGACGGTGGTTGCTCCGGTTGGGTGCT
GAATTAATCGGTACTTATGCAATTTCCGAATCTTTAGTTACTGAAAAATGGAATCTCTTA
GAGAGTAAGAGAGTGCTTTAATTTAGCTTAATTAGATTTATTTGGATTTCTTTTCACTATT
TGGATTGTAACTTTAGAATTTGTGTGTGTGTGTTGTTGCTTAGTCCTGAGATAAGATATAA
CATTAGCGACTGTGTATTATTATTACTGTCATTGTGTTATGTGAAACTTTGTTCTCTT
GTTGAAAAA

>G280 Amino Acid Sequence (domain in AA coordinates: 97-104,130-137-155-162,185-192)

MAFDLHHSASDTHSSELPSFSLPPYPQMIMEAIESLNDKNGCNKTTIAKHIESTQQTLP
PSHMTLLSYHLNQMKKTGQLIMVKNYMKPDPDAPPKRGRGRPPKQKTQAESDAAAAAVV
AATVVSTDPPRSRRPPKPKDPSEPPQEKVITGSGRPRGRPPKRPRTDSETVAPEPAAQ
ATGERRGRGRPPKVPTVVAVPGC*

>G3 (16..477)

GTTTGTCTTTTATCAATGGAAGAGAACAGAAGAGTCTACGATGAGAAAGAGAAGGCAG
CCACCTCAAGAAGAAGTGCCTAACCACGTGGCTACAAGGAAGCCGTACAGAGGGATACGG
AGGAGGAAGTGGGGCAAGTGGGTGGCTGAGATTCTGTGAGCCTAACAAACGCTACGGCTT
TGGCTTGGCTCTTTACACAACCGATATCGCCCGCTAGAGCCTACGACGTGGCCGCTTTC
TACCTCCGTGGCCCTCCGCACGTCTCAACTTCCCTGATCTTCTCTTGCAAGAAGAGGAC
CATCTCTCAGCCGCCACCACCGCTGACATGCCCGCAGCTCTTATAAGGGAAAAAGCGCG
GAGGTCCGGCGCCAGAGTCGACGCTCTTCTAGCTTCTGCCGCTCTTCGATGGCTCACTCC
ACTCCGCCGTAATAAAACCCGACTTGAATCAAATACCCGAATCCGGAGATATATAGTCA
ATTTATATACATGTAGTTTGTGTTTGTGATTAGAAAGATTACATTTACATACAAGATACA
CATAGATACTGGAAAATATAGGTATGTATACATTATATAAATTATCTTATGTATCAAAGAA
TTTTATAGATTCTGATTAGCTTTTTGTTTTGTTTTGATAAGAACTCTGATTAGTTGTC
CGGAGACAAAACCGCTAAGAGCAATCCATGAGAAGCTAGCGAGTGTTTTTTAGTTCAAG
TTGTAATATAAATGCATATTAATTCTTTAGTAATTTTGT

>G3 Amino Acid Sequence (domain in AA coordinates: 28-95)

MEREQEESTMRKRRQPPQEEVFNHVATRKPYPYRGIRRRKWKWVAEIREPNKRSRLWLGSY
TTDIAARAYDVAVFYLRGPSARLNFDPDLLLQEDHLSAATTADMPAALIREKAAEVGAR
VDALLASAAPSMHSTPPVIKPDNLNIPESGDI*

>G343 (1..795)

ATGGACGTCTATGGETTATCTTCACCAGACTTACTTCGAATCGACGACCTTCTTGATTTC
TCCAACGAAGACATCTTCTCCGCTTCTTCTCCGGTGGTTCCACCGCCGCTACTTCTCT
TCTTCTTTCCCTCTCTCTCAAAACCTAGTTTCCACCACCACCATCTCCCTTCTCCGCC
GATCATCACTCCTTCTCCACGACATTTGCGTTCACGATGATGACGACGCTCATCTTGAA
TGGCTTTTCGAATTCGTGGACGATTCTTTCGCTGATTTTCCGGCGAATCCATTAGGAGGA
ACTATGACTTCTGTCAAACTGAACTTCTTTCGGGGAAACCAAGAAGCAAACGATCA
AGAGCTCCTGCTCCTTTCGCCGGAACATGGTCTCCGATGCCACTGGAATCCGAGCATCAG
CAGCTTCACTCCGCCGCCAAATTCAGCCAAAGAAAGAACATCCGGCGGAGGAGGAGGA
GGAGGAGGAAGACATCAGTCATCGTCATCGGAGACTACGGAAGGAGGAGGAATGAGGAGA
TGTACTCACTGTGCATCGGAGAAAACGCCACAGTGGAGGACAGGACCACTTGGACCTAAA

ACACTATGTAACGCTTGTGGAGTCCGGTTTAAATCCGGTAGACTTGTACCGGAATATAGA
CCGGCTTCGAGTCCTACTTTTGTGTTTGAAGTCTCAAACTCTCACCAGAAAGTGATG
GAGCTTCGACGGCAGAAAGAAGTTATGAGACAACCACAACAAGTTCAACTTCATCACCAC
CACCACCGTTTTAG

>G343 Amino Acid Sequence (domain in AA coordinates: 178-214)
MDVYGLSSPDLLRIDLLDFSNEDIFSASSSGGSTAATSSSSFPFPQNPFSHHHLPSSA
DHHSFLHDICVPSDDAAHLEWLSQFVDDSFADFPANPLGGTMTSVKTETSFPGKPRSKRS
RAPAPFAGTWSMPLESEHQQLHSAKFKEQSGGGGGGGRRHQSSSEETTEGGGMRR
CTHCASEKTPQWRGTGLPKTLNACGVRFKSGRLVPEYRPASSPTFVLTHQHSNSHRKVM
ELRRQKEVMRQPQQVQLHHHHHPF*

>G363 (1..780)
ATGAGACCAATATTAGACCTCGAAATTGAAGCTTCATCGGGCAGTAGTAGCAGCCAAGTG
GCCTCAAACCTTGTCTCCGGTTGGGGAAGATTACAAACCAATCTCGCTGAATCTTAGCCTC
AGTTTCAACAACAACAACAATAATCTGGATCTTGAATCATCGTCTTTGACGCTGCCA
CTTCGAGCAGGAGTGAGAGTAGTAACCCGGAGCAGCAGCAGCAACAACAACCATCTGTA
TCAAAGAGAGTCTTCTCTGTAACTACTGCCAAAGGAAGTTCTATAGCTCTCAAGCGTA
GGTGGTCACCAAAACGCTCACAAACGTGAGAGAACACTCGCCAAACGCGCTATGCTATGG
GTCTTGTCTGGGTCTTCCCCGGTAGAGGATCAAGTAGCAATTATGCGGCTGCTGCCACAG
CAGCCGCTCTCGTGTGTCGCTTCACGGAAGCGGAAACGGGAACATGACATCGTTCAGG
ACTTTGGGAATCCGGGCACATTCTCGGCGCAGCAGCTCAGCATGACAAGGCAGACACCA
GAAACACTTATTAGAAACATTGCCAGGTTCAACCAGGGGTATTTCGGTAATTGTATACCT
TTTACGTGGAGGACGACGAGGCCGAGATGCTCTGGCCGGGGAGTTTCCGGCAAGCTACG
AATGCGGTTGCGGTTGAAGCGGGTAATGATAATTTAGGTGAAAGAAAAATGGATTTCTTG
GACGTCAAGCAAGCGATGGATATGGAAGTTCTCTTCCAGATCTAACCTTGAAGCTTTGA
>G363 Amino Acid Sequence (domain in AA coordinates: 87-108)
MRPILDLEIEASSGSSSSQVASNLSPVGEDYKPISLNLSLNFNNNNNNLDLESSSLTLP
LSSTSESSNPEQQQQQPSVSKRVFSCNYCQRKFYSSQALGGHQAHRERTLAKRAMLW
VLLGSSPVEDQVAIMRLLPQQLSCLPLHSGNGNMTSFRTLGIRAHSSAHDVSMTRQTP
ETLIRNIARFNQGYFGNCIPFYVEDDEAEMLWPGSFRQATNAVAVEAGNDNLGERKMDFL
DVKQAMDMESSLPDLTLKL*

>G370 (1..774)
ATGGACGAAACCAACGGACGAAGAGAAACTCAGGATTTTCATGAACGTCAACGTTGAATCC
TTCTCTCAGCTTCTTTTCATCCGCGTACTCTCCCAAAGAAAAGCCGCCATTATTCGT
CTCTTCGGCCAAGAGCTCGTTCGGTGATAACTCCGACAACCTTATCCGCAGAACCTTCTGAT
CATCAAACCATACCAAGAACGATGAGAGCTCTGAGAATATCAAGGACAAAGACAAAGAA
AAAGATAAGGACAAAGACAAAGATAACAACAACAGGAGATTTCGAGTGCTACTACTGC
TTCAGAAACTTCCCAACTTCTCAAGCCCTAGGTGGACATCAAAACGCTCACAACGTTGAA
CGTCAACACGCCAAACGCGGTTCCATGACATCATACCTTCATCATCATCAGCCTCATGAC
CCTCACCACATCTACGGCTTCTCTCAACAACCACCACCACCGTCACTATCCGTCTTGGACG
ACGGAAGCTAGATCATACTACGGCGGAGGGGACATCAAACGCGCTCGTACTACTCAAGG
AATACTCTTGCTCTCTCTTCTTAACCCACCGACAATCAACGGAAGTCCTTTAGGTTTG
TGGCGGTGACCGCCTTCCACGTCAACAAATACTATTCAAGGCGTTTACTCATCTTCACCA
GCTTCAGCGTTTAGGTTCGATGAGCAAGAGACTAATAAGGAGCCTAATACTGGCCGTAC
AGATTGATGAAACCAATGTGCAAGATCATGTGAGTCTCGATCTTCATCTCTGA

>G370 Amino Acid Sequence (domain in aa coordinates: 97-117)
MDETNGRRRETHDFMNVNVSFSQLPFIIRTPPKKAAIIRLFGQELVGDNSDNLSEPSD
HQTTTKNDESSENIKDKDEKDKDKDNNNNNRRFECHYCFRNFPTSQALGGHQAHRERE
RQHAKRGSMTSYLHHHQPHDPHHIYGLNNHHHRHYPSTTEARSYYGGGHHQTPSYYSR
NTLAPPSSNPPTINGSPLGLWRVPPSTSTNTIQGVYSSSPASAFRSHEQETNKEPNWPY
RLMKPNVQDHVSLDLHL*

>G385 (37..2202)
TAGGGTTTGCTTTCAGTTTCCGGAGTATAAGAAAAGATGTTTCGAGCCAAATATGCTGCTT
GCGGCTATGAACAACGCAGACAGCAATAACCACAACCTACAACCAGGAAGACAACAATAAT
GAAGGATTTCTTCGGGACGATGAATTCGACAGTCCGAATACTAAATCGGGAAGTGAGAAT
CAAGAAGGAGGATCAGGAAACGACCAAGATCCTCTTCATCCTAACAAGAAGAACGATAT
CATCGACACACCCAACCTTCAGATCCAGGAGATGGAAGCGTTCTTCAAGAGTGTCTCAC
CCAGATGACAAGCAAAGGAAACAGCTAAGCCGTGAATTGAATTTGGAACCTCTTCAGGTC

>G385 Amino Acid Sequence (domain in AA coordinates: 60-123)
MFEPNMLLAAMNNADSNHHNHNHEDNNNEGFLRDEDFDSPNTKSGSENGEGSGNDQDPL
HPNKKKRYHRHTQLQIQEMEAFFKECPHPDDKQQRKQLSRELNLPLEQVKFWFQNKRTQMK
NHHERHENSHLRAENEKLRNDNLRYREALANASCPNCGGPTAIGEMSFDEHQRLLENARL
REIDRISAIAAKYVGKPVSNYPLMSPPLPPRPLELAMGNIGGEAYGNPNPNDLLKSITA
PTESADKPIVIDLSVAAMEELRMVQVDEPLWKSALDEEYARTFPRGIGRPPAGRYSEA
PRESADVIMNHVNIVEILMDVNQWSTIFAGMVSRAMTLAVLSTGVAGNYNGALQVMSAEF
QVPSPLVPTRETYFARYCKQQGDGSWAVVDISLDSLQPNPPARCRRRASGCLIQELPNGY
SKVTWVEHVEVDDRGVHNLKYHVMVSTGHAFGAKRWVAILDRQCERLASVMATNISSGEVG
VITNQEGRRSMLKLAERMVISFCAGVSASTAHTWTTLSTGTGAEDVRVMTRKSVDDPGRSP
GIVLSAATSFWIPVPPKRVDFDLRDENSRNEWDILSNGGVVQEMAHIANGRDTGNCVSL
RVNSANSSQSNNMILQESCIDPTASFVIYAPVDIVAMNIVLNGGDPDYVALLPSPGFALP
DGNANSAGAPGGDGSLLTVAFQILVDSVPTAKLSLGSVATVMNLIAC TVERIKASMSCET
A*

TATAAATCTTCGTTTC TACTTTTTTTCTTCCATAATATAGTCAATTCTGTTTCTTAATT
AGGGCTCTTCTCTTGTGTTCTCCAATCTTTATTAGTTTATTTATTTATTTGGTTATTG
TATACAAATGGCAATGGCTTTAAACATGAATGCTTACGTAGACGAGTTCATGGAAGCTCT
TGAACCATT CATGAAGTAACTTCATCTTCTTCTACTTCGAATT CATCAAATCCAAACC
ATTAAC TCCTAATTTCA TCCCTAATAATGACCAAGTCTTACCGGTATCTAACC AAACCGG
TCCGATTGGGCTAAACCAGCTCACTCCAACACAAATCCCTCCAAATT CAGACAGAGTTACA
TCTCCGGCAAAACCAATCTCGTCGTCGCGCTGGTAGTCATCTTCTCACC GCTAAACCAAC
CTCAATGAAGAAAATCGAGGTAGCAACTAAACCGGTTAAACTATACCGAGGCGTAAGACA
GAGGCAATGGGGTAAATGGGTAGCTGAGATTCTGGCTACCTAAAAACCGAACCCGGTTATG
GCTCGTACGTTTCGAACGGCTCAAGAAGCTGCATTAGTTCATGCATCAAGCAGCTCATAA
GATCAGAGGAGACAACGCTCGTCTCAATTTCCCGAGACATTGTTTCGCTCAAGGACACTATAA

ACAGATATTGTCTCCGTCTATCAACGCAAAGATCGAATCCATCTGCAATAGTTCTGATCT
TCCACTGCCTCAGATCGAGAAACAGAACAAAACAGAGGAGGTGCTCTCTGGTTTTTCCAA
ACCGGAGAAAGAACCGBAATTTGGGGAGATATACGGATGCGGATACTCGGGCTCATCTCC
TGAGTCGGATATAACGTTGTTGGATTTCTCAAGCGACTGTGTGAAAGAAGATGAGAGTTT
CTTGATGGGTTTGCACAAGTATCCTTCTTTGGAGATTGATTGGGACGCTATAGAGAACT
CTTCTGAATCCATTTTATCTTTTGGATTCAATTTGTCTCTAAATTGTAGAATTTATTTTC
AGAGCTTTGTAAGGGAAGTTCTTGAATGAGAGTTGCAGAGGACTAGTGGAACTAACTCT
GTTTTCTTTTGTAAGTATTGTTTATAATGGGCCGTTGAATGGGCCTTATTGATTAAACA
GCCCAAGTTTTTAAAAAAAAAAAAAAAAAAAAA

>G439 Amino Acid Sequence (domain in AA coordinates: 110-177)
MAMALNMNAYVDEFMEALEPFMKVTSSTSNSSNPPLTPNFIPNNDQVLPVSNQTGPI
GLNQLTPTQILQIQTELHLRQNSRRRAGSHLLTAKPTSMKKIDVATKPKLYRGVRQRQ
WGKWVAEIRLPKNRTRLWLGTFFETAQEAALAYDQAAHKIRGDNARLNFPDIVRQGHYKQI
LSPSINAKIESICNSSDLPLPQIEKQNKTEEVLSGFSKPEKEPEFGEIYCGYSGSSPES
DITLLDFSSDCVKEDESFLMGLHKYPSLEIDWDAIEKLF*

>G440 (237..1301)
AAAAAATCACTGTTTCATAACACGTTTTTCTCTCTACCCACCAAAAAAATCTTTTGT
TCTTGTGTACCAAAAAATCTCGTGATAAATCTCTTCAAACTTTGTTTTATTTCTTCTGA
TTCTCTCGAAATCTCTCTCAACAAACCCAGAACTTTCTTGTATTGCAAGCTTTTCTTC
CTTTTATATTCTTCATTTTGTATGCGAATATAGAGAGAGTCCATAAAGAAACAGTAATGG
ACGAATATATTGATTTCGACCATTGAAGTACACAGAGCACAAGACTTCAATGACTAAAT
ACACCAAAAAGTCATCGGAAAACTTTCCGGTGGTAAGTCATTGAAAAAGGTTAGTATTT
GTTATACTGATCCTGACGCAACAGATTCACTCAAGTGACGAAGACGAAGAAGATTTCTGT
TTCTCTCGCCGGAGAGTCAAAAGATTCTGTTAACGAGATCACTGTTGAGCCTAGCTGTAACA
ACGTCGTCACCGGAGTTTCGATGAAAGATAGAAAGAGACTCTCTTCTCTCCGATGAAA
CTCAATCTCCGGCTCGAGTCGTCACGTCCTAATAACAAAGTTTCAGTCTCCGGTCAGA
TAAAGAAAGTTCCGTGGTGTAGACAACGGCCATGGGGGAAATGGGCGCGGAGATTAGAG
ATCCGGAGCAACGTCGGAGGATTTGGCTCGGGACTTTTGAGACGGCGGAGGAAGCTGCCG
TGGTTTATGATAACGCCGCTATAAGACTCCGTGGACCGGACGCTTTAACTAATTTCTCCA
TACCGCCTCAAGAAGAGGAAGAAGAAGAAGAACCGGAACCGGTTATTGAGGAGAAACCGG
TTATTATGACGACGCCAACCAACAACATCGAGTTCTGAATCAACTGAAGAAGATTTAC
AACATCTCTCATCTCTTCTCGGTTCTCAATCACCGGTCAGAAGAGATTCAACAAGTAC
ACAACCGTTTAAATCAGCTAAACCCGAACCGGGGTTTCAAATGCACCATGGTGGCATA
CCGGGTTTAAATACCGGTTTAGGTGAATCAGACGATTCAATTTCTTTGGATACTCCGTTTC
TTGACAACTATTTCATGAATCACCACCAGAGATGTCAATATTTGACCAACCAATGGATC
AAATTTCTGTGAAAATGATGATATCTTCAATGATATGTTGTTCTTGGGTGGTGAAACTA
TGAACATTGAAGATGAGTTAACAAAGTTCTAGTATCAAAGATATGGGTTCAACGTTTAGTG
ATTTTGTATGATTCAATGATATCAGATCTATTAGTTGCTTAATATGATGATGAGAGTGAAG
AAGAAACCATCAAGCAATATCTATGGTGTGACTGAAAAATTTTGGTGTACTTTTTTTT
CTTTCATAAGTTCATGAGCTTTTTTGTCTTTTTTTTTTAATAATTTATTAGTTTGTCA
GGAGCTTGTAACAGTTTTGGAGAAATAGTGGAATAATAGTTTAATTAATAAAAAAAAAA
AAAAAA

>G440 Amino Acid Sequence (domain in AA coordinates: 122-189)
MDEYIDFRPLKYTEHKTSMTKYTKSSEKLSGGKSLKKSICYPDPDATDSSSEDEEDF
LFPRRRVKRFVNEITVEPSCNNVTVGSMKDRKRLSSSDETQSPASSRQRPNNKVSVSG
QIKKFRGVRQRPWKWAAEIRDPEQRRRIWLGTFFETAEEAAVVYDNAAIRLRGPDALTNF
SIPPQEEEEEEPEPVIEKPVIMTPTPTSSSESTEEDLQHLSSPTSVLNHRSEEIQQ
VQQPFKSAKPEPGVSNAPWWHTGFNTGLGESDDSFPLDTPFLDNYPNESPPEMSIFDQPM
DQIFCENDDIFNDMLFLGGETMNIIEDELTSSSIKDMGSTFSDFFDLSLIDLLVA*

>G5 (417..1421)
TTTTTTTTTGTCAATCTCCCCCTAATCTGTTGTTTCTCGCTTCTTCTTCTGTTAATCATC
TGTCCTTTCAAAAAGAAAGAAAAAGAAAAATTCGATTTCTGGGTTTGTTTTTGTGCATACA
GAAAAAATCAAGCTTATGAATTTGTGTTAATTTTTTGTTTTAAATTTGAAAGGCAGGTT
TTTTCAGAACGAGATCGTTTTTCAAATTTCTTCTGATTTTACCTCTTTTTTTCTTCTTA
GATTTTAGTGAATCGAGGGTGAAATTTTTGATTCCTCTTTTCGGATCTACACAGAGGTT
GCTTATTTCAAACCTTTTAGATCCATTTTTTTTAAATTTCTCGGAAAAATCCCTGTTTC
TTTACTTTTTTATAAGTCTCAGGTTCAATTTTTTTCGGATTCAAATTTTATTTTAAATGG

CAGCTGCTATGAATTTGTACACTTGTAGCAGATCGTTTCAAGACTCTGGTGGTGAAGTCA
TGGACGCGCTTGTACCTTTTATCAAAAGCGTTTCCGATTCTCCTTCTTCTTCTTGCAG
CGTCTGCGTCTGCGTTTCTTACCCCTCTGCGTTTCTCTCCCTCCTCTCCCCGGTTATT
ACCCGGATTCAACGTTCTTGACCAACCGTTTTCATACGGGTCGGATCTTCAACAAACCG
GGTCATTAATCGGACTCAACAACCTCTCTTCTCTCAGATCCACCAGATCCAGTCTCAGA
TCCATCATCTCTTCTCCTCCGACGCATCACAACAACAACCTCTTCTCGAATCTTCTCA
GCCAAAGCCGTTACTGATGAAGCAATCTGGAGTCGCTGGATCTTGTTCGCTTACGGTT
CAGGTGTTCTTTCGAAGCCGACGAAGCTTTACAGAGGTGTGAGGCAACGTCAGTGGGAA
AATGGGTGGCTGAGATCCGTTTGCCGAGAAATCGGACTCGTCTCTGGCTTGGGACTTTTG
ACACGGCGGAGGAAGCTGCGTTGGCCTATGATAAGGCGGCGTACAAGCTGCGCGGCGATT
TCGCCCCGGCTTAACCTCCCTAACCTACGTCATAACGGATTTACATCGGAGGCGATTTCG
GTGAATATAAACCTCTTCACTCCTCAGTCGACGCTAAGCTTGAAGCTATTGTAAAAGCA
TGGCGGAGACTCAGAAACAGGACAAATCGACGAAATCATCGAAGAAACGTGAGAAGAAGG
TTTCGTCGCCAGATCTATCGGAGAAAGTGAAGGCGGAGGAGAATTCGGTTTCGATCGGTG
GATCTCCACCGGTGACGGAGTTTGAAGAGTCCACCGCTGGATCTTCGCCGTTGTCGGACT
TGACGTTTCGCTGACCCGGAGGAGCCGCCGAGTGAACGAGACGTTCTCGTTGGAGAAGT
ATCCGTCGTACGAGATCGATTGGGATTTCGATTCTAGCTTAGGGGCAAAATAGGAAATTC
GCCGCTTGCAATGGAGTTTGTGAAATTCGATGACTGGCCCAAGAGTAATTAATTAAT
ATGGATTAGTGTAAATTTTCGATGTTAATATTTGTATTATGGTTTGTATTAGTCTCTCT
GTGTCGGTCCAGCTTTCGGGTTTTCGAGCTCGACCATGCCACAGTTTTTCATTTTATG
TAATCTTTTTTCTTTGTCTTATGTAATTTGTAGCTTCAGTTTCTTCATCTATAATGCA
ATTTTATTATGATTATGTG

>G5 Amino Acid Sequence (domain in AA coordinates: 149-216)
MAAAMNLYTCSRFSQDSGGLMDALVPFIKSVSDSPSSSSAASASAFSLHPSAFSLPPLPG
YYPDSTFLTQPSYSGDLQQTGSLIGLNNLSSSQIHQIQSQIHHPLPPTHNNNNNSFNL
LSPKPLLKMQSGVAGSCFAYGSGVPSKPTKLYRGVRQRHWGKVAEIRLPRNRTRLWLGT
FDTAEEAALAYDKAAYKLRGDFARLNFNLRHNGFHIGGDFGEYKPLHSSVDKLEAICK
SMAETQKQDKSTKSSKKREKKVSSPDLEKVKAEENSVSIGGSPPVTEFEESTAGSSPLS
DLTFADPEEPPQWNFTFSLEKYPSEYIDWDSILA*

>G550 (1..1374)

ATGGCTGATCCGGCGATTAAGCTCTTTGGAAAGACGATTCCCTTTACCTGAGCTTGGTGT
GTTGATTCTTCTTAGCTATACCGGATTTTAAACGAAACTCAGATTCCGTGTTCCGGTTA
TCAGATTCCGTGTACCGGCGATGATGATGATGAAGAGATGGGTGATTCCGGTTTAGGACGA
GAAGAAGGTGATGATGTTGGTGATGGTGGAGGAGAGAGCGAGACTGATAAAAAGGAAGAA
AAAGATAGTGAGTGTGAGGAAGAGTCATTGAGGAATGAATCTAATGATGTTACTACTACT
ACATCGGTATAACTGAAAAACGGAACAACAAGAGCTGCAAGACGAATGAAGAGTCA
GGTGGTACTGCTTGTCTCAAGAGGGGAAGTTAAAGAAACCTGATAAGATTCTACCGTGT
CCGCGATGTAACAGCATGGAACCAAGTTCTGTTACTACAACAACATAATGTTAACC
CCTCGCCATTTCTGCAAGAAATGTCAGAGATATTGGACAGCTGGTGGAACGATGAGGAAT
GTTCCGGTTGGTGTGAGGAGACGTAAGAATAAGAGTCCAGCTTCTCATTATAACCGTCAT
GTAAGTATAACATCTGCGGAAGCTATGCAGAAGGTGGCGAGAAGTGAATCTTCAACATCCT
AATGGTGCAATCTTCTCACTTTTGGCTCTGATTCTGTGCTTTGTGAATCTATGGCTTCT
GGATTGAATCTTGTGAGAAGTCATTGTTGAAGACACAACTGTATTGCAAGAACCCAAT
GAAGGCTTGAAGATTACGGTTCCGTTAAACCAGACAAACGAAGAAGCTGGAACAGTCAGC
CCGTTACCAAAAAGTTCCATGCTTTCCAGGACCACCACCAACTTGGCCTTACGCTTGGAAC
GGAGTTTCGTGGACGATTTTACCGTTTACCTCCACCGGCTTACTGGAGCTGCCCCGGG
GTTTCACCGGGGGCATGGAACAGCTTCACATGGATGCCACAACCAATTACCATCTGGT
TCCAATCCAAATCTCTACACTAGGTAAACATTCACGTGACGAGAACGCTGCTGAACCA
GGAACCGCTTTTGTGAAACCGAGTCACTTGGTAGGGAGAAAAGCAAAACCGAGAGATGC
TTGTGGGTTCCCAAGACGCTGAGGATTGATGATCCAGAGGAAGCTGCTAAAAGTTCCATC
TGGGAAACATTAGGGATCAAAAAGACGAAATGCGGATACTTTCGGAGCTTTCAGATCA
TCAACCAAGAAAAAGCAGTCTTCTGAAGGAAGACTTCCGGAAGAAGACCGGAGTTG
CAAGCGAATCCTGCTGCTCTTCTAGGTCAGCAAACTTCCATGAGAGCTCATAG

>G550 Amino Acid Sequence (domain in AA coordinates: 134-180)
MADPAIKLPGKTIPLPELGVDSSSYTGFLTETQIPVRLSDSCTGDDDDDEEMGDSGLGR
EEGDDVGDGGGESETDKKEEKDSECQESLRNESNDVTTTSGITEKTETTKAAKTNEES
GGTACSQEGKLKPKDILPCPRCNSMETKFCYNNYNNVNPQRHFCKKCQRYWTAGGTMRN

>G670 (28..1152)

>G670 Amino Acid Sequence (domain in AA coordinates: 14-122)

>G760 (175..1878)

179

CCACTCAACGATGGTCTGTACCTGGAAACGAATGATCTCAGCTGTGCTCAACAAGATGAT
TTTAACTTCGAAGATTATCTCAGCTTCTTTGATGATGAGGGTTTGACTTTTGACGATTCT
CTATTAATGGGACCTGAAGATTTTCTTCCCAACCAAGAAGCCCTTGACCAGAAACCTGCC
CCTAAAGAATTGGAGAAGGAGGTCGCAGGAGGCAAGAGGCAGTGGAGGAAAAGGAAAGT
GGCGAAGGATCTTCTTCAAAAACAGAGCCACATGCTTGGAGCCATTCTACTCCATCTTCATT
TACCCATTTCTCAAAAAGACGAGCCACATGCTTGGAGCCATTCTACTCCATCTTCATT
GCTTCACAGTTCCAAACAAAGGACGCAATGCGTCTACACGCAGCACAATCTTCTGGTTCA
GTTACAGTGACTGCAGGTATGATGAGAATATCAAACATGACTCTAGCAGCGGACAGCGGT
ATGGGCTGGTCATATGACAAGAACGGTAACCTCAACGTAGTCCTTTTCATTTCGGGGTAGTC
CAACAGGATGATGCGATGACTGCCTCGGGAAGCAAGACAGGAATTACGGCGACAAGAGCT
ATGTTAGTCTTCATGTGTTTATGGGTTCTCTACTCTCTGTTAGCTTCAAAATAGTAACC
ATGGTGTCTGCTCGGTAATAGGATCAAAGTTGAATCGTCTCAAAGACTTTTTTTGGTGT
TGTACCTCTCCAATCATATAGCCTTTAACTTTGGCAGTGCTTTGCTGCTCAATATTTAAA
TTTTAAAAA

>G760 Amino Acid Sequence (domain in AA coordinates: 12-156)
MGRGSVTS LAPGRFRHPTDEELVRYLKRKVCNKPFDKDAISVTDIYKSEPWDLPDKSKL
KSRDLEWYFFSMLDKKYSNGSKTNRATEKGYWKTGKDREIRNGSRVVGMMKTLVYHKGR
APRGERTNWMHEYRLSDEDLKKAGVPQEAYVLCRIFQKSGTGPKNGEQYGAPYLEEWE
EDGMTYVPAQDAFSEGLALNDDVYVDIDIDEKPENLVVYDAVPILPNYCHGESSNNVES
GNYSDSGNYIQPGNNVVDGGYFEQPIETFEEDRKPIIREGSIQPCSLFPEEQIGCGVQD
ENVVNLESSNNNVFVADTCYSIDIPIDHNYLPDEPFMDPNNNLPLNDGLYLETNDLSAQQ
DDFNFEYDLSFFDDEGLTFDDSLMGPEDFLPNQEAALDQKPAPKELEKEVAGGKEAVEEK
ESGEGSSSKQDTFKDFDSAPKYPFLKKTSHMLGAIPTPSSFASQFQTKDAMRLHAAQSS
GSVHVVTAGMMRISNMTLAADSGMGWSYDKNGNLNVVLSFGVVQDDAMTASGSKTGITAT
RAMLVFMCLWVLLLSVSFKIVTMVSAR*

>G831 (92..1987)

TTCTTTTCATCGTTGTGTCTATTATAAATATATGTCAATTTGGTTTCTAAAAAATCTACC
ATTGATTGATTGATTTTTTTTTCTTTAAGAGATGAATTTATTACAAGAATCTCATCTCG
GACTAAGAAGGCCAATCTTTACTACGTAACCTAGTTGCTCTTCTCTGCATCGCTAGCTA
CCTTCTCGGTATTTGGCAAAACACGGCGGTTAATCCACGCGCCGCTTCGATGATTGAGA
CGGTACACCGTGCGAGGGATTACACGACCTAATTCTACGAAAGATCTCGACTTCGACGC
GCATCACAACATTCAAGATCCACCTCCGGTGACGGAACCGCCGTTAGTTTCCCGTCGTG
TGCCGCCGCGTTGAGCGAGCACACGCCATGCGAAGACGCGAAGCGATCGTTGAAATCTC
GAGGGAGAGATTGGAGTATAGGCAAAGGCATTGTCCCGAGAGAGAAGAAATCTTGAAGTG
CAGAATTCGGCGCCGTACGGTTACAAAACGCGTTCCGATGGCCGCGAGTCGTGACGT
GGCGTGGTTTCGCTAATGTGCCTCACACGGAGCTTACGGTTGAGAAAAAGAAATCAGAATTG
GGTCCGGTACGAGAATGATCGGTTTTGGTTCCCTGGTGGAGGTACGATGTTTCCACGTGG
CGCTGATGCTTACATTGATGATATCGGACGGTTGATTGATCTCAGCGACGGCTCTATCCG
TACAGCCATCGATACCGTTGCGGGGTGGCTAGCTTCGGTGATATCTTTTATCAAGAAA
CATTACAACGATGTCTATTGTCACCAAGAGACACACAGAACTCAAGTCCAGTTCGCACT
CGAGCGTGGTGTGCGCGCGATGATCGGAATCATGGCTACAATCCGCTACCGTACCCCTC
TAGAGCCTTTGATTTAGCACATTGCTCTCGTTGCCTTATTCGGTGGGGCCAAAACGATGG
GGCTTACTTGATGGAGGTGGATAGGGTTTTAAGACCAGGAGGGTACTGGATACTTTCTGG
ACCGCCGATTAATTGGCAGAAACGGTGGAAAGGGTGGGAACGGACCATGGATGATTTGAA
TGCAGAGCAGACTCAGATCGAGCAGGTCGCGAGAAGCTTGTGTTGGAAGAAAGTTGTTCA
AAGAGATGATCTTGCTATTTGGCAAAAACCTTTAACCACATTGACTGTAAGAAAACCG
AGAGGTTTTGAAAAATCCGGAGTTTTGTCTCATGATCAAGATCCCGACATGGCCTGGTA
TACGAAGATGGATTGTTGTTTACACCATTAACCTGAAGTTGATGACGCTGAGGATCTAAA
GACGGTGGCCCGGAGGGAAGGTAGAAAAGTGGCCGGCTAGATTAAACGCGATTCTCCGAG
AGTAAACAAAGGCGCTCTCGAGGAAATCACACCTGAAGCTTCTTGGAGAACACGAACT
GTGGAAACAGAGAGTTTCTTATTACAAGAAGTTAGATTACCAAGTTGGGTGAAACCGGGAG
ATACAGAAACTTAGTCGACATGAACGCTTACCTCGGTGGATTTCGCGCGGGCTCTAGCGGA
TGATCCGGTCTGGGTCAATGAACGTTGTCCCGGTGAGGCTAAGCTCAATACGCTCGGTGT
CATCTACGAGCGTGGTCTAATCGGAACGTATCAAACTGGTGTGAAGCCATGTGACGTA
TCCAAGAACGTATGATTTTATCCATGCTGACTCGGTTTTTACATTTGTACCAAGGTCAATG
TGAACCGGAGGAGATATTGTTGGAGATGGACCGAATTTCTAGACCGGGTGGTGGTGTGAT
TATAAGAGATGACGTGGACGTTTTGATCAAGGTTAAGGAATTAACCAAAGGATTAGAATG

GGAAGGTAGAATTGCTGACCACGAGAAGGGTCCTCATGAAAGAGAGAAGATTTACTATGC
GGTGAAACAGTATTGGACCGTTCTCTGCGCCTGATGAAGATAAAACAACACTAGTGCTCT
CTCCTGATTTTTGAGTTTTTTTTTTTCTTACAATGTTTTTTTTTTTTTTTCAATTTTT
TATACAACAATAAATTCTCAATAATTGTTGTGTCGCGGCCG

>G831 Amino Acid Sequence (domain in AA coordinates: 470-591)
MNLFRISRRTKKANLYVTVLVALLCIASYLLGIWQNTAVNPRAAFDDSDGTPCEGFTRP
NSTKDLDFDAHNIQDPPPVTEAVSFPSCAAALSEHTPCEDAKRSLKFSRERLEYRQRH
CPEREELKCRIPAPYGYKTPFRWPASRDVAVFANVPHTELTVEKKNQNWVRYENDRFWF
PGGGTMTFPRGADAYIDDIGRLIDLSDGSIRTAIDTGCQVASFAYLLSRNITMTSFAPRD
THEAQVQFALERGVAMPIMIGIMATIRLPYPSRAFDLAHCSRCLIPWQNDGAYLMEVDRVL
RPGGYWILSGPPINWQKRKKGWERTMDDLNAEQTIQEQVARSLCWKKVVQRDDLAIWQKP
FNHIDCKKTREVLKNPEFCRHDQDPMAWYTKMDSCLTPLPEVDDAEDLKTAVAGGKVEKW
PARLNAIPPRVNKGALBEITPEAFLENTKLWKQVRVSYYKKLDYQLGETGRYRNLDVMNAY
LGGFAAALADDPVWVMNVVPEAKLNTLGVYIERGLIGTYQNWCEAMSTYPRTYDFIHAD
SVFTLYQGQCEPEEILLEMDRILRPGGGVIIRDVDVLIKVKELTKGLEWEGRIADHEKG
PHEREKIYYAVKQYWTVPAPDEDKNNTSALS*

>G864 (503..1534)
TGCAAAAACATTTTCTGTCTCTCCTCTGCCCAAATTTTTTTTCTTTCCAGGAATATTTTC
CTAGAAAAACCCAAGCAAAGCTTTAACCCTCTCCTCTCCAAAGTAGCATCTTCCTCTT
TTTCTATTTCTCCTTTCTCTTCTTATCTCTCTCGTTTGTGAACGATTCTTAAGAAT
ATAACCAAAGCCCTTTTCTCCTTTCTTCAACTTTCCGGGAAAAATCTTCACGCAGCAAG
GTTTCTCTCTCGGCTCTCGCAGTGTTCCTCGGCCTTTTGTCTTTCTATAAAAAA
TTCCGCTCCTTTAAGAAACTTTTTCCACCTAGAGAAGAAGAAGATCACTCTTGTG
TTCAAGTTCTCTCTTTAATAAAAAATCCATCTTTATTCTTTGTCTCTTTCTTTTGC
TTTCCCTAATCTCTATGTTATAAACACACAGAGAGAAACAAAGTCACAGTCTCGAGTCAA
AAACAGAGAATAACGAAAGAAAAATGGAAGCGGAGAGAAAGAAATGGTTCTACCGAGAATCAA
ATTACAGAGACACAAACCAACACGACAACAATCGTATCGGAGTTAACCAACACTCACCA
AACCAGGATTCTTCGTATCTCAGTCACTGACCCAGACGCTACTGATTCTCCAGTGACGA
CGAAGAAGAAGAACATCAACGCTTTGTCTCTAAACGCCGCTCGTGTAAAGAGTTTGTCAA
CGAAGTCTATCTCGATTCCGGTGCTGTTGTTACTGGTAGTTGTGGTCAAATGGAGTGCAG
GAAGAGACAAAAGAGAGCGGTTAAATCGGAGTCTACTGTTTCTCCGGTTGTTTCAGCGAC
GACGACTACGACGGGAGAGAAGAAGTTCCGAGGAGTGAGACAGCGTCCATGGGGAAAAATG
GGCGGCGGAGATAAGAGATCCGTTGAAACGTGTACGGCTCTGGTTAGGTACTTACAACAC
GGCGGAAGAAGCTGCTATGGTTTACGATAACGCCGCTATTTCAGCTTCGTGGTCCCAGCGC
TCTGACTAATTTCTCAGTCACTCCGACAACAGCGACGGAGAGAAAGCCCCACCACCGTC
TCCGGTGAAGAAGAAGAAGAAGAAAAACAACAAAGCAAAAAATCCGTTACTGCTTCTTC
CTCCATCAGCAGAAGCAGCAGCAACGATTGTCTCTGCTCTCCGGTGTCTGTTCTCCGATC
TCCTTTCCGCGTCGACGAATTCTCCGGCATTCTTCATCACCAGTCGCGGCCGTTGTAGT
CAAGGAAGAGCCATCCATGACAACGGTATCTGAAACTTTCTCTGATTTCTCGGCGCCCTT
GTTCTCAGATGATGACGTGTTTCGATTTCCGGAGCTCAGTGGTTCCCGACTATCTCGGCGG
CGATTTATTTGGGGAAGATCTATTCACGGCGGATATGTGTACGGATATGAACCTCGGATT
CGATTTCCGATCCGGATTATCCAGCTGGCACATGGAGGACCATTTTCAAGATATCGGGGA
TCTATTCCGGTCCGATCCTCTTTTAGCTGTTTAATAATATTTTAAATAAATAAATAGTTA
TACCGGCCGTTACTAAACGGAACCGGAGAAAGTTTTGTATACCGGTGACATAAAATCTCG
GTTATGTTTCGTAATCTTTTTTTCTTTGTTATATATAAAAAATGAATGAACTGAATTA
TGTAAGTTAATGGTGATAATTATTAACGTTTAAAGTTTTGAAAAAAAAAAAAAAAAAAAA
AAAAAA

>G864 Amino Acid Sequence (domain in AA coordinates: 119-186)
MEAEKRMVLPRIKFTEHKNTTTIVSELNTHQTRILRISVTDPDATDSSSDDEEEHQR
FVSKRRRVKFKFVNEVYLDGSAVVTGSCGQMESKKRQKRAVKSESTVSPVVSATTTTTGEK
KFRGVRQRPWGKWAAEIRDPLKRVRLWLGTYNTEAEEAMVYDNAAIQLRGPDALTNFSVT
PTTATEKKAPPPSPVKKKKKKNKSKSVTASSSISRSSNDCLSPVSVLRSPFAVDEF
SGISSSPVAVVVKEEPSMTTVSETFSDFSAPLFSDDDVDFRSSVVPDYLGDLFGEDL
FTADMCTDMNFGDFGSGLSWHMEDHFQDIGDLFGSDPLLAV*

>G884 (31..1575)
TTTTTTTTTGTTTGTAAATTTTGGGGATCGATGTCGGAAGGAAGAAGCTCCGTCGACA
TCGAAGTCCACCGGAGCTCCGTCGCGTCCGACTTTATCTCTCTCCACGGCCGTTTAGT

GAGATGTTCTTTAAACGGTGGCGTTGGATTTCAGTCCTGGTCCGATGACTCTGGTCTCTAAT
ATGTTCCCTGATTCCGATGAGTTTAGGTCTTTCTCTCAGCTTCTCGCTGGAGCCATGTCT
TCTCCAGCGACTGCAGCTGCTGCTGCTGCTGCTGCGACGGCTAGTGATTACCAGAGACTT
GGTGAAGGGACTAATAGCTCTAGTGGTGATGTTGACCCGAGATTCAAGCAAAACAGACCA
ACCGGTTTGATGATTTCTCAATCTCAATCGCCGTCGATGTTACCGTACCGCCTGGTTTA
AGTCCAGCTATGTTGCTCGATTACCAAGCTTTTTGGGTCTTTTCTCTCCCGTTTCAGGGA
TCATATGGAATGACACATCAGCAAGCTCTAGCTCAAGTCACTGCTCAAGCAGTTCAAGCC
AATGCCAATATGCAACCACAAACAGAGTACCCTCCTCCCTCTCAAGTTCAATCATTTTCA
TCGGGTCAAGCGCAGATCCCGACCTCGGCTCCACTACCAGCTCAAAGAGAAACCTCAGAT
GTAACCATCATAGAGCACAGGTACCAACAGCCTCTAAATGTTGACAAACCAGCTGATGAT
GGCTATAACTGGCGAAAATATGGGCAAAAGCAAGTTAAAGGTAGCGAGTTTCCACGAAGC
TATTACAAGTGTACTAATCCAGGATGTCTGTCAAGAAGAAGGTTGAGAGATCTCTTGAT
GGACAAGTAACGGAGATTATCTACAAAGGTCAGCACAAATCATGAACCTCCTCAAAACACT
AAGCGAGGTAACAAAGATAACACCGCGAATATAAATGGGAGTTCGATAAATAACAATCGC
GGGAGTTCTGAATTGGGGGCATCAGTTTCAAACCTAATAGCTCCAACAAGACTAAGAGA
GAGCAACATGAAGCAGTAAGTCAAGCTACGACAACAGAGCACTTGTCTGAGGCAAGTGAC
GGTGAAGAAGTTGGTAATGGAGAAACTGATGTGAGAGAGAAAGATGAGAATGAGCCTGAT
CCCAAGAGAAGTACAGAAAGTTCCGATTTCAGAACAGCTCCTGCTGCTTCACATAGA
ACTGTGACAGAGCCTAGAATTATTGTCCAAACGACGAGTGAAGTTGATCTTCTAGATGAT
GGATATAGGTGGCGTAAATATGGACAGAAAGTTGTCAAAGGGAATCCTTATCCGAGGAGC
TACTACAAGTGCACAACACCAGGATGTGGTGTGAGGAAACATGTAGAGAGAGCAGCAACA
GATCCAAAAGCTGTAGTAACAACATATGAAGGAAAACATAACCATGACCTTCCCGCTGCT
AAATCAAGCAGCCATGCCGCTGCAGCGGCACAGTTAAGGCCAGATAATCGACCTGGCGGT
TTGGCTAACTTAAATCAACAGCAGCAGCAACAGCCCGTTGCGCGGCTAAGGCTTAAAGAA
GAGCAAACTTGAAGAGAAGAACTCTTGACCGTTTTTCATTACAAAAGCTTTCAAAT
TCCACTCACACACTGTCTGAAAATCTAGCAGTTTGCAGGAAAGAAACAGCTTCAAGAG
GTTGTAGTTCTCTATGTTCTGGTGTAAACCTTAAAGCTTTTTAGGGTTTTTCAGATTTT
TGTTTACTAATACTGTATGTGAATCTTTTGTACATGAGGAAGAAAATTACAGGGGGATA
TTTTGTGTTGTATCTTTTGTGTTATGTTTCAGTAAAAGATAGGTCTTACATTTTGTGTA
AAAAAAAAAAAAAAAAAAAA

>G884 Amino Acid Sequence (conserved domain in AA coordinates:227-285, 407-465)

MSEKEEAPSTSKSTGAPSRPTLSLPPRPFSEMFNGGVGFSPGPMTLVSNMFPDSDEFRS
FSQLLAGAMSSPATAAAAAAATASDYQRLGEGTNSSSGDVPDFKQNRPTGLMISQSQS
PSMFTVPPGLSPAMLLDSPSFLGLFSPVQGSYGMTHQQLAQVTAQAVQANANMQPQTEY
PPPSQVQSFSSGQAQIPTSAPLPAQRETS DVTIIIEHRSQQPLNVDKPADDDGYNWRKYGQK
QVKGSEFPRSYKCTNP GCPVKKKVERSLDGQVTEIIYKGQHNHEPPQNTKRNJNDNTAN
INGSSINNNRGSSELGASQFQTNSSNKTREQHEAVSQATTEHLSEASDGEEVNGNETD
VREKDENEPPDKRRSTEVRISEPAPAASHRTVTEPRIIVQTTSEVDLLDDGYRWRKYGQK
VVKGNPYPRSYKCTTPGCGVRKHVERAATDPKAVVTTYEGKHNHDLPAKSSSHAAAAA
QLRPDNRPGGLANLNQQQQPVARLRLKEEQTT*

>G898 (161..772)

GAAAAAAGATTCAAAAACCTAGATTTCAAAAAATCGATTGGCTGTCAAATTTCTCTCC
GGCGATTTTCTCGAGTGAAATTCGGCTCAAGGTGATTATAGCGATCATCGAATCAAATT
GATTGAAGAGGTACAAAGGTTAGTTACTTTGAGCTGAAAGATGAACACGTGAGAGGTGAG
AGTACCTCGAGGAAATCGACGAGGAAAGCTGTGATTGATCTGAATGCGGTACCTGTTGA
TCAAGAAGGGACCTCTGCTTCTGTTAGAACTCTTACGGTGCCTATTACACCGTCTCAGCC
TGCTCCTACGATGATTGATGTCGATGCTATTGAGGATGATGTTATTGAATCATCCGCTAG
TGCTTTTGTGTAAGCTAAAAGCAAATCAAGAAATGCACGTCGGAGACCTTTGATGGTTGA
TGTAGAGTCAGGAGGTACGACTAGATTCCCTGCCAACATAAGCAACAAACGAGAAGGAT
TCCTTCTAGTGAATCTGTCTGACTGTGAGCATGCCTCTGTAAATGATGAAGTCAACAT
GTCTTCGAGAGTGTCTAGATCAAAGGCTCCAGCTCCTCCACCAGAAGAGCCAAAGTTTAC
ATGTCCAATCTGCATGTGTCCCTTTACGGAGGAGATGTCAACCAAGTGCGGTACATCTT
CTGCAAGGGATGTATAAAGATGGCAATATCTCGCCAGGGCAAATGCCCTACTTGTAGGAA
AAAGGTTACTGCAAAAGAGCTGATTTCGAGTTTTCTTCCAACCACTAGATGAGTGGTCCG
GCAACATCACCAGCCACCCTGTCTAATGGTTTATCAGACTATCCTCCTATTCACTTTGGA
ACATTGAAGGGACTTCGTTGACTTGGTATTTTTGAATATTTTGCTTTGTTGGAAGAGAAA
TATTTCAGTGATCAAGAAGCCAGAAGGCCCTATCATTTCGATGGATATCATTGGTAATAACT

CTTTGTTTTTAGTTGTTGTTCTATGTAATTTAGGTCTCTGCAAACCTCTCAGTCGATACT
CTTCTCTCTTGATAGATGATAAGATATATGGAAAAAAATTAATATTGAATCTTTACTA
AAA

>G898 Amino Acid Sequence (domain in AA coordinates: 148-185)
MNTSEVRVPRGNRRRKAVIDLNAVVDQEGTSASVRLTVPIPSQPAPTMIDVDAIEDD
VIESSASAFEAESKSRNARRRPLMVDVESGGTTRFPANISNKRRIIPSSSEVIDCEHAS
VNDEVNMSSRVSRKAPAPPPEEPKFTCPICMCPFTEEMSTKCGHIFCKGCIKMAISRQG
KCPTCRKKVTAKELIRVFLPTTR*

>G900 (1..648)
ATGGGGAAGAAGAAGTGCAGTTATGTTGTGGTGTAGCGAGAATGTATTGTGAGTCAGAT
CAAGCGAGTTTATGTTGGGATTGTGACGGTAAAGTTCACGGAGCTAATTTTCTGGTGGCG
AAACACATGCGTTGTCTTCTATGTAGCGCGTGTCTAGTCACACACGCCCTTGGAAAGCTTCT
GGGCTGAATCTTGGCCCAACTGTTTCTATCTGTGAGTCTTGTTTAGCTCGTAAGAAGAAT
AACAACAGCTCCCTCGCCGGGAGGGATCAGAATCTTAACCAAGAAGAAGAGATCATTGGT
TGTAACGACGGAGCTGAGTCTTATGATGAGGAAAGCGATGAGGATGAAGAAGAAGAAGAA
GTGGAGAATCAGGTTGTTCCGGCTGCGGTGGAGCAAGAACTTCCGGTGGTGAGTTCGTCTG
TCTTCGGTTAGTAGTGGTGAAGGAGATCAGGTGGTGAAGAGGACGAGACTTGATTTGGAT
CTTAACCTCTCCGATGAGGAGAACCAATCTAGACCATTGAAAAGATTATCGAGAGACGAA
GGTTTGTCAAGATCAACTGTTGTGATGAATAGCTCAATCGTGAAATTACACGGAGGGAGG
AGAAAAGCAGAGGGATGTGATACATCATCGTCTGCTTCTGTTTTATTGA

>G900 Amino Acid Sequence (domain in AA coordinates: 6-28, 48-74)
MGKKKCELCGGVARMYCESDQASLCWDCDGKVVHGANFLVAKHMRCLLCSACQSHTPWKAS
GLNLGPTVSICESCLARKKNMNSLAGRDQNLNQEEIEIGCNDGAESYDEESEDDEEEEEE
VENQVVPAAVEQELPVVSSSSSVSSGEGDQVVKRRLDLNLSDEENQSRPLKRLSRDE
GLSRSTVVMNSSIVKLHGRRRKAEGCDTSSSSSPY*

>G913 (108..806)
CATTCAAAAACATCATATATATACACAAACACACTTTGATACAACAAAAAACAGAAC
ACAAACAAAAACACATTGTAACATTAGTTTAAGCATTAAGCTTCTTTATGTGCAATAATA
ATAATTCTCCGACCACCGTGAATCAAGAAACGACGACGTCTCGTGAAGTCTCAATCACAT
TGCCTACTGATCAATCTCCTCAAACCTCACCAGGATCATCTTCTTCTCCTTCACCGAGAC
CTTCCGGTGGATCACCAGCGAGAAGAACGGCGACTGGATTATCCGGCAAGCACTCTATTT
TCAGGGGGATTTCGACTACGTAACGGAAATGGGTATCGGAGATTAGAGAGCCACGTAAAA
CGACAAGAATTTGGCTCGGACTTATCCGGTACCGGAGATGGCTGCCGCCGCTTACGACG
TGGCTGCGTTAGCTTTAAAAGGACCCGACGCCGTTTTGAATTTTCTGGTTTAGCTTTGA
CTTACGTGGCTCCGTTTCAAACCTCTGCTGCGGATATAAGAGCGGCTGCTAGTAGAGCAG
CGGAGATGAAGCAACCGGATCAGGGTGGGGATGAGAAGGTATTGGAACCGGTTCAACCCG
GCAAAGAGGAAGAATTAGAAGAAGTGTCGTGTAACCTCGTGTTCGTTGGAGTTTATGGATG
AGGAAGCGATGTTGAATATGCCGACTTGTGTGACGGAGATGGCTGAAGGGATGTTGATGA
GTCCACCGAGAATGATGATACATCCGACGATGGAAGATGATTCCGCCGAGAATCATGAAG
GAGATAATCTTTGGAGTTATAAATGAATCCATTGAAGCTGCTCTCTTTTTTTATTGTTTTT
CGGTGCAATGAGATTTTCCCCCTTTTTTTTTTTCTTTTTGGGTCGCTGTT

>G913 Amino Acid Sequence (domain in AA coordinates: 62-128)
MSNNNSPTTVNQETTSREVSITLPTDQSPQTSPPSSSSPSRPSGGSPARRTATGLSG
KHSIFRGIIRLNRNGKWVSEIREPRKTRIWLGTYPVPEMAAAAYDVAALALKGPDAVLNFP
GLALTYVAPVNSAADIRAAASRAEMKQPDQGGDEKVLPEVPQPGKEEELEEVSCNSCSL
EFMDEEAMLNMPDLLTEMAEGLMSPPRMMIHPTMEDDSPENHEGDNLSYK*

>G937 (45..1046)
TGGAAAAAGTTTGAATTTTTTAATTCGAATCGAGAAAAATAAAAAATGGGTTCTTTAGGTG
ATGAGCTTAGTTTGGGATCGATCTTTGGGAGAGGAGTTTCGATGAATGTTGTGGCGGTTG
AGAAAGTTGATGAACATGTTAAGAAGCTTGAAGAAGAGAAGAGAAAGCTCGAAAGTTGTC
AACTTGAGCTTCTCTGTCTTTGCAGATTTTAAACGATGCGATTTGTATCTGAAGGATA
AGAGATGTTTCAGAGATGGAGACTCAACCATTGTTGAAAGATTTCATTTCTGTTAATAAAC
CTATTCAAGGAGAAAGAGGAATAGAATTGCTGAAAAGAGAGGAGCTAATGAGGGAGAAGA
AGTTTCAGCAATGGAAGCTAATGATGATCACACTAGTAAGATCAAGAGCAAGCTTGAGA
TTAAGAGAAATGAGGAGAAATCTCTATGTTGTTGATTCCAAAGGTGGAACTGGTTTAG
GCCTCGGTTTAAAGTTTCGAGTTTCGATAAGAAGAAAGGGATTGTTGCCTCATGTGGCTTTA
CTTCTAACTCTATGCCACAACCAACACAGCAGTACCACAACAACAGCATTCTTTA

AGCAGCAAGCTTTACGGAAGCAAAGAAGGTGTTGGAATCCAGAGTTGCATCGCCGATTTG
TCGATGCATTGCAACAGCTAGGTGGACCGGGAGTGGCAACTCCTAAACAAATTAGAGAAC
ATATGCAAGAAGAAGGCTTAACCAATGATGAAGTCAAGAGTCATTTACAGAAATACAGGT
TACACATCAGGAAGCCAAATTCGAATGCGGAGAAACAATCAGCAGTTGTTTATAGGGTTTA
ACTTGTGGAATTTCTTCAGCACAGATGAAGAAGAGACATGTGAAGGAGGAGAATCATTGA
AGAGAAGCAATGCGCAATCAGATTCTCCTCAAGGTCCTTTGCAGTTACCGTCTACAACAA
CAACAACCTGGTGGAGATAGTAGCATGGAAGATGTTGAAGATGCTAAGTCTGAGAGCTTTC
AACTGGAGAGATTGAGATCACCATAAATCTCAAGAAACCAACTCTTGATCACGGTTTTG
TTATTTTGGATTCTACTATATCTATTAGTAGTGAATGAGAACAATAATTATAGAAAGG
TTTATAGATATATATATAGAGAAAAAGAGAGAGTGGGATGGTTCAAATTATTTGCAGA

>G937 Amino Acid Sequence (conserved domain in AA coordinates:197-246)

MGSLGDELSLGSIFGRGVSMNVVAVEKVDHVKKLEEEKRKLESCQLELPLSLQILNDAL
LYLKDKRCSEMETQPLLKDFISVKNPIQGERGIELLKREELMREKKFQQWKANDDHTSKI
KSKLEIKRNEEKSPMLLIPKVETGLGLGLSSSIRRKIVASCFTSNMPQPPTPAVPQ
QPAFLKQOALRKQRRCWNPFLHRRFVDALQQLGGPGVATPKQIREHMQEEGLTNDEVKSH
LQKYRLHIRKPNNSNAEKQSAVVLGFNLWNSSAQDEEETCEGGESLKRSNAQSDSPQGPLQ
LPSTTTTTGGDSSMEDVEDAKSESFQLERLRSP*

>G960 (63..1538)

TACCGTCGACCCACGCGTCCGAGTGATTCAAAGTCGGAAAGAAACCCTAAAGAAGAGGA
TTATGGGTGCTGTATCGATGGAGTGCCTTCTTTAGGTTTCAGATTCAGACCTACCGATG
AAGAGCTCGTCAATCACTACCTCCGTCTCAAGATCAACGGACGTCCTCCGATGTCCTG
TCATCCCTGATATCGATGTCTGCAAATGGGAACCTTGGGATCTTCTGCTCTCTCGGTGA
TTAAGACGGATGATCCAGAGTGGTCTTTTTCTGCCCTCGTGATCGGAAATACCCTAATG
GTCATCGCTCTAACAGAGCAACTGACTCTGGCTATTGGAAAGCTACTGGTAAAGATCGTA
GCATCAAGTCTAAGAGAGACTTTAATCGGTATGAAGAAGACTCTTGCTCTTCTATCGTGGAC
GAGCTCCTAAAGGTGAGCGGACTAATTGGATTATGCACGAGTATCGTCCCACTCTTAAGG
ATCTTGATGGCACTTCCCCTGGCCAAAGCCCTTACGTTCTTTGTGCGCTCTTCCACAAGC
CTGATGATCGGGTTAATGGTGTCAAGTCCGATGAAGCAGCTTTTACGGCCAGCAACAAAT
ACTCACCTGATGATACATCATCTGATCTTGTTCAAGAAACACCTTCTCTGATGCTGCTG
TTGAGAAACCATCAGATTATTCAGGTGGATGCGGTTATGCTCATAGTAATAGTACCGCAG
ATGGGACAATGATTGAGGCACCTGAAGAGAATCTTTGGTTATCTTGACCTTGAAGATC
AAAAGGCACCACTACCGTGTATGGATTCTATATATGCTGGTGATTTCAGTTACGATGAGA
TTGGATTCCAAATTTCAAGATGGTACCAGCGAACCAGATGTATCACTAACAGAATTGTTGG
AGGAGGTGTTCAATAACCCTGATGACTTCTCTTGCGAGGAATCGATCAGTCGAGAGAATC
CAGCAGTCTCACCAATGGGATATTTTCTGCTAAAATGCTGCAGTCTGCAGCACCAG
AGGATGCTTTCTTCAACGACTTCATGGCTTTCACTGATACAGATGCTGAGATGGCGCAAT
TGCAGTATGGTTTCAGAAAGGTGGAGCTTCTGGTTGGCCAAGTGACACTAATTCATACTATA
GTGATTTGGTTTCAGCAAGAGCAAATGATCAATCATAACACAGAGAACAACCTCACAGAAG
GGAGAGGGATAAAGATCCGGGCTCGACAGCCTCAGAACCGGCAGAGTACAGGATTGATAA
ACCAGGGTATTGCTCCAAGGAGAATCCGTCTGCAGCTGCAGTCTAACTCTGAAGTAAAG
AACGAGAGGAGGTGAATGAAGGACACACTGTTATTCCCGAGGCCAAAGAAGCTGCAGCTA
AATACTCAGAGAAGAGTGGTTCTTTGGTTAAACCTCAAATAAAGCTCAGGGCGCGGGGAA
CTATAGGCCAAGTAAAGGAGAGAGATTTGCAGACGACGAGGTACAGGTGCAGAGCACAA
AGAGAGAGAGAGAGAGAATCAAATGTAGTTTAAATGTAATTAGGGATGATGCAATGTTAGC
ATGTTTGTGTGTTGTAACCTTAAAACTTATTTAGGAATCTGATAAAAGTTACTGTTGAAA
AAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>G960 Amino Acid Sequence (domain in AA coordinates: 13-156)

MGAVSMESLPLGFRFRPTDEELVNHYLRRLKINGRHSVDRVIPDIDVCKWEPWDLPALSVI
KTDDPEWFFFCPRDRKYPNGHRNDRATDSGYWKATGKDRSIKSKKTLIGMKKTLVFYRGR
APKERTNWINMHEYRPTLKDLDGTSFGQSPYVLCRLFHKPDDRVRNGVKSDEAFTASNKY
SPDDTSSDLVQETPSSDAAVEKPSDYSGGCGYAHNSNADGTMIEAPEENLWLSCDLEDQ
KAPLPCMDSIYAGDFSDEIGFQFQDGTSEPDVSLTELEEVFNPNPDDFSCEESISREN
AVSPNGIFSSAKMLQSAAPEDAFNDFMAFTDTDAEMAQLQYSGEGGASGWPSDTNSYYS
DLVQQEQMINHNTENNLTTEGRGIKIRARQPQNRQSTGLINQGIAPRRIRLQLQSNSEVKE
REEVNEGHTVIPEAKEAAKYSEKSGSLVKPQIKLRARGTIGQVKGERFADDEVQVQSTK
RERERIKCSLM*

>G991 (6..533)

GAAAAATGGAAGAAGAAAAGAGATTGGAGCTAAGGCTAGCTCCTTGTACCAATTCA
 CTTCCAACAACAACATCAATGGATCTAAACAAAAAGCTCGACCAAGAAACATCATTCC
 TTTCCAATAACAGGGTTGAGGTAGCTCCAGTGGTGGGATGGCCCGGTGAGATCATCCC
 GGAGAAACCTAACCGCACAACTAAAGGAGGAGATGAAGAAGAAGGAGAGTGATGAAGAGA
 AGGAATTGTACGTTAAGATCAACATGGAAGGAGTTCCAATAGGAAGAAAAGTCAACCTTT
 CAGCTTATAACAACCTACCAACAGCTTTCACATGCCGTTGACCAACTCTTCTCTAAGAAAAG
 ATTCGTGGGATCTAAACAGACAATACACTTTGGTCTACGAAGACACTGAAGGAGATAAAG
 TTCTGGTCGGGGATGTTCTTGGGAGATGTTTGTATCTACTGTAAAGAGGTTGCATGTTT
 TAAAGACCTCCACGCCTTCTCACTCTCACCTAGAAAACATGGCAAGGAATAGAGAGAGG
 TTGGCCAAAACATCAGTTCGATGGTTTGTTTTAAATGTAATTTTGTGGAACTAATGG
 GGTTTGGCTTTTGATTTACTGGTTTTCTTTTCACTTATGTACTAGGTTTTTGTCTGCTAT
 GTTATTTCTTGTTTTGGTTGTAAATATGCTGTTCTGTTTAAAGAAATCGGGGGTTAGTATGT
 TATCGTGTGTATAAAAAATAGTGTAAAGCACGTAAGTTGATTACAAAAAATAAAAAA
 AAAAAA

>G991 Amino Acid Sequence (domain in AA coordinates: 7-14,48-59,82-115,128-164)

MEEKRLELRLAPPCHQFTSNNNINGSKQKSSTKETSFLSNRVEVAPVVGWPPVRSSRR
 NLTAQLKEEMKKKESDEEKELYVKINMEGVPIGRKVNLSAYNNYQQLSHAVDQLFSKKDS
 WDLNRQYTLVYEDTEGDKVLVGDVPWEMFVSTVKRLHVLKTSFAFSLSPRKHGKE*

>G748 (98..1444)

CCACGCGTCCGCACTCTCCCAATCTCTCTTTTAAACAACAAAAAATCAGAGAGA
 CATAGAGAGAAGAAGACGGAACAGAGGCTCCAAAAAATGATGATGGAGACTAGAGATCC
 AGCTATTAAGCTTTTCGGTATGAAAATCCCTTTTCCGTCGGTTTTTGAATCGGCAGTTAC
 GGTGGAGGATGACGAAGAAGATGACTGGAGCGGCGGAGATGACAAATCACCAGAGAAGGT
 AACTCCAGAGTTATCAGATAAGAACAACAACAACTGTAACGACAACAGTTTTTAACAATTC
 GAAACCCGAAACCTTGGACAAAGAGGAAGCGACATCAACTGATCAGATAGAGAGTAGTGA
 CACGCTGAGGATAATCAGCAGACGACCTGATGGTAAAACCTAAAGAAACCGACTAA
 GATTCTACCGTGTCCGAGATGCAAAAGCATGGAGACCAAGTTCGTATTACAACAATA
 CAACATAAACACGCTCGTCATTTCTGCAAGGCTTGTGAGAGATATGGGACTGCTGGAGG
 GACTATGAGGAATGTTCTGTGGGGGCGAGACGTCGTAAGAACAAGCTCATCTTCTCA
 TTACCGTCACATCACTATTTCCGAGGCTCTTGAGGCTGCGAGGCTTGACCCGGGCTTACA
 GGCAACACAAGGGTCTTGAGTTTTTGGTCTCGAAGCTCAGCAGCAGCACGTTGCTGCTCC
 CATGACACCTGTTATGAAGCTACAAGAAGATCAAAAGGTCTCAAACGGTGCTAGGAACAG
 GTTTCACGGGTTAGCGGATCAACGGCTTGTAGCTCGGGTAGAGAAATGGAGATGATTGCTC
 AAGCGGATCCTCTGTGACCACCTCTAACAATCACTCAGTGGATGAATCAAGAGCACAAAG
 CGGCAGTGTGTTGAAGCACAAATGAACAACAACAACAATAACATGAATGGTTATGC
 TTGATCTCCAGGTGTTCCATGGCCTTACACGTGGAATCCAGCGATGCCCTCCACAGGTTT
 TTACCCGCTCCAGGTTATCCAATGCCGTTTACCCTTACTGGACCAATCCCAATGCTACC
 ACCGCATCAATCCTCATCGCTATAAGCCAAAAGTGTTCAAATACAACTCTCCGACTCT
 CGGAAAGCATCCGAGAGATGAAGGATCATCGAAAAGGACAATGAGACAGAGCGAAAACA
 GAAGGCCGGGTGCGTTCTGCTCCGAAAACGTTGAGAATAGATGATCCTAACGAAGCAGC
 AAAGAGCTCGATATGGACAACATTGGGAATCAAGAACGAGGCGATGTGCAAAGCCGGTGG
 TATGTTCAAAGGGTTTGATCATAAGACAAAGATGTATAACAACGACAAAGCTGAGAACTC
 CCCTGTTCTTTCTGCTAACCTGCTGCTCTATCAAGATCACACAATTTCCATGAACAGAT
 TTAGAGTTACATATGTATATATATATATGATGATTGATGTATGATAGATGATACTGG
 AGAATGATGAGTTTTTTGAGAATCAAACCTTTTCTTTCTAGTGATTGCCTTTATTCC
 TTTACATGTTTTTGGTTCTCTGTACACTATTTGATTACCTTTTTTACTTTCTTTCTTCAT
 TTGTCAGGAAATGTTGGAAGATAACATTAATGGTAAAAAGTTGGTGTGGACCGTTGTTGC
 GTTGGCATTTCAAAAAATAAAAAA

>G748 Amino Acid Sequence (domain in AA coordinates: 112-140)

MMMETRDPAILKFLGMKIPFPVVFESAVTVEDEDDWSGGDDKSPEKVTPELSDKNNNNC
 NDNSFNNSKPELTLDKEEATSTDQIESDTPEDNQQTTPDGKTLKKPTKILPCPRCKSMET
 KFCYNNYNINQPRHFCKACQRYWTAGGTMNRNVPVGAGRRKNKSSSSHYRHITISEALEA
 ARLDPLGLQANTRVLSFGLEAQQHVAAAPMTPVMKLQEDQKVSNGARNRFGHLADQRLVAR
 VENGDDCSSGSSVTTSSNNHVSDESRAQSGSVVEAQMNNNNNNNMNGYACIPGVWPYPTWN
 PAMPPPGFYPPPGYPMPFPYPTWIPMLPPHQSSSPISQKCSNTNSPTLGKHPRDEGSSKK
 DNETERKQKAGCVLPKTLRIDDPNEAAKSSIWTTLGIKNEAMCKAGGMFKGFDHKTKMY
 NNDKAENSPVLSANPAALSRSHNFHEQI*

>G247 (1..660)

ATGAGAATGACAAGAGATGGAAAAGAACATGAATACAAGAAAGGTTTATGGACAGTGGAA
GAAGACAAGATCCTCATGGATTATGTCCGAACATCATGGCCAGGGCCACTGGAACCGCATC
GCCAAGAAAACCTGGGCTCAAGAGATGTGGGAAAAGCTGTAGGTTGAGATGGATGAACATC
TTAAGCCCTAATGTTAACAGAGGCAATTTTACTGACCAAGAAGAAGATCTCATCATCAGA
CTCCACAAGCTCCTCGGCAACAGATGGTCGTTGATAGCGAAAAGAGTTCCGGGAAGAACA
GACAACCAAGTAAAGAATTACTGGAACACACATCTCAGCAAGAACTTGGTCTCGGAGAT
CATTCAACTGCCGTCAAAGCCGCATGCGGTGTAGAGTCTCCACCGTCTATGGCCCTTATA
ACCACAACGTCTCTCTCATCAAGAGATCTCCGGTGGAAAAAATTCAACTCTAAGGTTT
GACACTTTAGTTGACGAATCCAACTCAAACCAAATCCAACTAGTCCACGCAACACCA
ACTGACGTAGAAGTTGCAGCTACGGTTCCAAATCTGTTGATACCTTTTGGGTTCTTGAA
GACGACTTCGAGCTTAGTTCACTCACTATGATGGATTTTACTAATGGGTATTGCCTTTGA
>G247 Amino Acid Sequence (domain in AA coordinates: 15-116)
MRMTRDGKEHEYKGLWTV EEDKILMDYVRTHGQGHWNRIAKKTGLKRCGKSCRLRWMNY
LSPNVNVRGNFTDQEEDLIIRLHKLLGNRWSLIAKRVPGRTDNQVKNYWNTHLSKKLGLGD
HSTAVKAACGVESPSPMALITTTSSSHQEISGGKNSTLRFDTLVDESKLKP KSKLVHATP
TDVEVAATVPNLFDTFWVLEDDFELSSLTMMDF TNGYCL*

>G585 (111..2039)

CTCTCAAACATTTCTCTGTTTGTTCGGCGAAAACGGCAACTGTTTCATCAAATGACAAA
CACAAAAACCTTAACATCTAGTTTGTATCCTCTCTGATACCTCAAAAAAATGGATGAAG
AAACAATGGCTACCGGACAAAACAGAACAACTGTGCCAGAGAATCTGAAGAAACACCTCG
CAGTTTCAGTTGCAAAACATTCATGGAGTTATGGTATCTTTTGGTCTGTCTCTGCTTCTC
AGTCTGGAGTTT TAGAATGGGGAGATGGATACTATAATGGAGATATCAAAAACGAGGAAGA
CGATTCAAGCTTCGGAGATCAAAGCTGATCAGCTTGGTCTACGGAGGAGCGAGCAGCTTA
GCGAGCTTTACGAGTCTCTCTCGTCTGATCTTCTTCTTCAAGCGGTGCTGCCGGAT
CTCAAGTCACCAAGACGAGCTTCCGCCGCCGACCTTTCACCGGAAGATCTCGCCGACACCG
AGTGGTACTATTTGGTTTGTATGTCTTTCGTCTTCAACATTGGTGAAGGAATGCCTGGAC
GGACGTTTGC AAAACGGTGAACCGATATGGTTGTGCAACGCTCATACGGCGGATAGTAAAG
TGTTTAGCCGTCTCTTCTAGCAAAAAGTGCTGCGGTTAAGACAGTGGTTTGCTTCCCGT
TCCTTGGAGGAGTCGTTGAGATTGGTACCACAGAACATATTACGGAAGACATGAATGTAA
TACAATGCGTGAAGACATCATTCCTCGAAGCCCCGTGATCCGTACGCTACAATATTACCAG
CAAGATCCGATTATCACATCGACAACGTTCTTGATCCGCAACAGATTCTAGGCGACGAGA
TTTACGCGCCTATGTTCAAGTACGGAGCCTTTTCCAACAGCTTCTCCGAGCAGAACTACCA
ACGGTTTCGATCAAGAACATGAACAAGTAGCAGATGATCATGATTCTTTTCATGACCGAAA
GAATCACTGGAGGAGCTTCTCAGGTGCAAAAGCTGGCAGCTCATGGACGACGAGCTTAGTA
ACTGCGTTACACAGTCGCTAAATTCCAGCGATTGCGTCTCTCAAACGTTTGTGTAAGGGG
CGGCTGGACGGGTTGCTTACGGTGCAAGAAAGAGTAGAGTTCAAAGACTAGGGCAAATTC
AAGAGCAACAGAGAAATGTGAAGACATTGTCAATTTGATCCAAGAAACGACGACGTTCAAT
ACCAAAGTGTGATCTCAACGATTTTTAAGACCAACCATCAGTTAATTCTCGGACCGCAGT
TTCGAAACTGCGATAAACAGTCAAGCTTCACTAGGTGGAAGAAATCATCGTCATCATCAT
CAGGAACCGCCACGGTCACGGCACCATCACAGGAATGTTAAAGAAAATTATTTTCGATG
TTCCGCGAGTGCACCGAAGAGAGTTAATGTTGGACTCACCAGAAGCCAGAGATGAAA
CTGGGAACCATGCGGTTT TAGAGAAGAAGCGCCGAGAAATTGAACGAACGGTTTCATGA
CCTTGAGAAAAATCATTCCTGCAATCAACAAGATCGATAAAGTATCGATTCTTGACGATA
CGATAGAGTATCTTCAAGAACTCGAGAGACGGGTTCAAGAACTAGAACTTGCAGAGAAT
CAACCGATACAGAGACTCGTGGGACGATGACGATGAAGAGGAAGAAACCATGCGACGCAG
GAGAAAGAACATCAGCTAATTGCGCAAATAATGAACAGGAAATGGGAAGAAGGTGTCGG
TTAACAATGTTGGTGAAGCCGAGCCAGCAGATACCGGTTTACTGGTTTAAACCGATAATT
TAAGGATCGGTTCTGTTTGGTAATGAGGTGGTTATTGAGCTTAGATGTGCTTGAGAGAAG
GAGTATTGCTTGAGATAATGGATGTGATTAGTGATCTCCATTGGATTCTCATTCGGTTC
AATCCTCGACCCGAGACGGTTTGCTCTGCTTAACCGTCAATTGCAAGCACAAAGGGTCAA
AAATAGCGACACCAGGAATGATCAAAGAAGCACTTCAAAGGGTTCATGGATCTGTTGAA
GACTACTTAGTTAAAATTGACAGCAAAGAAAAAACATTCCCGGTTTGGTTTCTATTCTTT
GGTTTTCTTCTAACCGGGTTTTAGGAATTAATGTTATGTTTATCATTTGTTTTTGT
TTTTTGTGCTTTTTTTCCGTTGCTTAACGTAGGTGAAGAGGAACATACTATGCGTA
TTTTGTTGAGGTAGATTATTTAAGGTATTAGTAATAGTAATAGCCAGTTTAGATGAT
TTTGTGTTCTTTTGTGTT

>G585 Amino Acid Sequence (domain in AA coordinates: 436-501)
MDEETMATGQNRTTVPENLKKHLAVSVRNQWSYGFWSVSASQSGVLEWGDGYNGDIK
TRKTIQASEIKADQLGLRRSEQLSELYESLSVAESSSSGVAAGSQVTRRASAAALSPEDL
ADTEWYYLVCMFVFNIGEGMPGRTFANGEPWLNCNAHTADSKVFSRSLAKSAAVKTIV
CFPFLGGVVEIGTTEHTEDMNVIQCVKTSFLEAPDPYATILPARSDYHIDNVLDPPQIL
GDEIYAPMFSTPEFPPTASPSRTTNGFDQEHQVADDHDSFMTERTGGASQVQSWQLMDD
ELSNVCVHQLNSSDCVSQTVEGAAGRVAYGARKSRVQRLGQIQEQQRNVKTLSDPRND
DVHYQSVISTIFKTNHQLILGPQFRNCDKQSSFTRWKSSSSSSSGTATVTAPSQGLKKI
IFDVPRVHQKEKLMDSPEARDETGNHAVLEKKRREKLNRFMTLRKIIPSINKIDKVISI
LDDTIEYLQELERRVQELSCRESTDTETRGTMKRRKPCDAGERTSANCANNETGNGK
KVSVMNVGEAEPADTGFTGLTDNLRIQSGFNEVVIELRCAWREGVLLIMDVISDLHLDS
HSVQSSTGDGLLCLTVNCKHKGSKIATPGMIKEALQVAVIC*

>G634 (1..798)
ATGGAGCAAGGAGGAGGTGGTGGTGGTAATGAAGTTGTGGAGGAAGCTTACCTATTAGT
TCAAGACCTCCTGCTAACAACCTTAGAAGAGCTTATGAGATTCTCAGCCGCCGCGGATGAC
GGTGGATTAGGAGGTGGAGGTGGAGGAGGAGGAGGAAGTGCTTCTTCTTCATCGGGA
AATCGATGGCCGAGAGAAGAACTTTAGCTCTTCTTCGGATCCGATCCGATATGGATTCT
ACTTTTCGTGATGCTACTCTCAAAGCTCCTCTTTGGGAACATGTTTCCAGGAAGCTATTG
GAGTTAGGTTACAAACGAAGTTCAAAGAAATGCAAAGAGAAATTCGAAAACGTTACAGAA
TATTACAAACGTACTAAAGAACTCGCGGTGGTCTGTCATGATGGTAAAGCTTACAAGTTC
TTCTCTCAGCTTGAAGCTCTCAACTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
CATCAACCAAGAACAAACAACAACAACAACAAGAGATGGTCTAGAGCTCGGAA
CAATCATCATTTACCATCATCATCAAGATGGCCAAAGGCAGAGATTCTAGCGCTTATAAAC
CTGAGAAGTGGAATGGAACCAAGGTACCAAGATAATGTACCTAAAGGACTTCTATGGGAA
GAGATCTCAACTTCAATGAAGAGAATGGGATACAACAGAAACGCTAAGAGATGTAAAGAG
AAATGGGAAAACATAAACAATACTACAAGAAAGTTAAAGAAAGCAACAACAGCAACTAC
ACAACAGAATCAATGA
>G634 Amino Acid Sequence (domain in aa coordinates: 62-147, 189-245)
MEQGGGGGGNEVVEEASPISSRPANNLEELMRFSAAADDDGLGGGGGGGGGSSSSSSG
NRWPREETLALLRIRSDMDSTFRDATLKAPLWEHVSRLKLELGYKRSKKCKEKFENVQK
YYKRTKETRGGRHDGKAYKFFSQLEALNTTPPPPPSHPHAHQPEKQQQQPQQEMVMSSE
QSSLPSSSSRWPKAEILALINLRSGMEPRYQDNVPGKLLWEEISTSMKRMGYNRNAKRCKE
KWENINKYYKKVKESNNSNYMKNQ*

>G676 (1..612)
atgagaaagaaagtaagtagtagtggtgacgaaggaacaatgagtacaagaaaggttg
tggacagtagaagaagacaaaatcctcatggattatgtcaaagctcatggcaaaggtcac
tggaaatcgatttgccaaaaagactggtttaaagagatgtggaaagagttgtagattgagg
tggatgaattatctcagccctaattgtgaaaagaggcaatttcaccgagcaagaagaggat
cttatcattaggtccacaagttgcttggttaattaggtggtctttaaattgctaaaagagtg
ccgggtcgaaacggataatcaagtgagaactattggaacacgcatcttagtaagaaactc
ggaatcaaagatcagaaaaccaaacagagcaatggtgatattgtttatcaaataatctc
ccgaatcctaccgaaacatcagaagaacgaaaatctcgaatattgtcgataacaataat
atcctcggagatgaaattcaagaagatcatcaaggaagtaactacttgagttcactttgg
gttcatgaggatgagtttgagcttagcacactcaccaacatgatggactttatagatgga
cactgtttttga

>G676 Amino Acid Sequence (domain in AA coordinates: 17-119)
MRKKVSSSGDEGNNEYKKGLWTVEEDKILMDYVKAHGKGHWNRIAKKTGLKRCGKSCRLR
WMNYLSPNVKRGNFTEQEEDLIIRLHKLLGNRWSLIAKRVPGRTDNQVKYWNTHLSKKL
GIKDQKTKQSNQDIVYQINLPNPTETSEETKISNIVDNNILGDEIQEDHQGSNYLSSLW
VHEDEFELSTLTNMMDFIDGHCF*

>G682 (1..228)
ATGGATAACCATCGCAGGACTAAGCAACCCAAGACCAACTCCATCGTTACTTCTTCTTCT
GAAGAAGTGAGTAGTCTTGAGTGGGAAGTTGTGAACATGAGTCAAGAAGAAGAAGATTG
GTCTCTCGAATGCATAAGCTTGTCGGTGACAGGTGGGAAGTATAGCTGGGAGGATCCCA
GGAAGAACCCTGGAGAAATTGAGAGGTTTTGGGTGATGAAAAATTGA

>G682 Amino Acid Sequence (domain in AA coordinates 27-63)
MDNHRRTKQPKTNSIVTSSSEEVSSLEWVNVMSQEEEDLVSRMHKLVGDRWELIAGRIIP

GRTAGEIERFWVMKN*

>G635 (1..993)

ATGGAGATCATGCGTCCAGGGGTCCTCAGAAAACACTTTGAAAGGAAAAATAAGAATCACA
ACGCGGTGCATGTGGCTTGACAAAGGAAGACTTTTAGATGCACTTCACAAAGCAGCTCAT
GCTGCTCTATCAAGTTGTCTCTGTGACATGTCCCTTGTCTCACATGGAAAGAACAGTCTCC
GAAGTCCTGAGGAAGATTGTAAGGAAGTACAGTGGTAAAAGGCCTGAAGTCATCGCTATA
GCCACTGAGAATCCAATGGCTGTCCGAGCTGATGAGGTCAGTGCAGACTGTCTGGTGAT
CCAAGTGTGGTTCTGGAGTTGCAGCTTTAAGGAAAGTTGTTGAAGGAAATGACAAAAGA
AGTCGGGCGAAGAAAGCACCTTCACAAGAAGCTTCCCCCAAAGAAGTAGATCGCACTTTG
GAAGATGATATCATTGATAGTGCAAGACTACTGGCTGAAGAAGAAACTGCGGCATCAACA
TACACGGAAGAAGTTGATACGCCCGTTGGGAGTTCTTCAGAAGAGTCAGACGATTTTGG
AAATCATTATCAATCCATCATCGTCACCTTCACCGAGTGAAACAGAAAATATGAATAAG
GTAGCTGATACGGAGCCTAAAGCAGAGGGTAAGGAAAACAGCAGAGACGACGATGAATTA
GCTGATGCTTCAGATTCTGAAACCAAGTCATCACCAAAACGTGTGAGGAAGAACAATGG
AAACCGGAGGAGATAAAGAAGGTAATCAGAATGCGAGGAGAGCTGCACAGTAGATTTCAA
GTGGTGAAAGGTAGAAATGGCATTGTGGGAAGAGATCTCTTCAAATCTATCAGCTGAAGGA
ATCAATCGAAGCCCGGACAAATGCAAAATCTCTCTGGGCATCACTTATTTCAGAAATACGAG
GAGAGCAAGGCTGATGAGAGAAGCAAGACGAGTTGGCCACATTTTGAGGATATGAACAAC
ATTTTGTGAGAGCTAGGCACACCTGCGTCTTAA

>G635 Amino Acid Sequence (domain in AA coordinates: 239-323)

MEIMRPGVSENTLKGKIRITTRCMWLDKGRLLDALHKAHAALSSCPVTCPLSHMERTVS
EVLRKIVRKYSGRKPEVIALATENPMAVRADEVSAARLSGDPVSGVAAALRKVVEGNDKR
SRAKKAPSQEASPKVEDRTLEDDIIDSARLLAEEETAASTYTEVDTPVGSSEESDDFW
KSFINPSSSPSPSETENMNKVADTEPKAEGKENSRRDDELADASDSETKSSPKRVRKNKW
KPEEIKKVIRMRGELHSRFQVVKGRMALWEEISSNLSAEGINRSPGQCKSLWASLIQKYE
ESKADERSKTSWPHFEDMNNILSELGTFAS*

>G1068 (150..1310)

GAGAGTTGTTAGCTAGCTCACACGCTTTCGCTTAAAACTCAAAAACCTGCACCTTCTCGT
CTATTTTCTCGGCATTTCGTAACAGAAAAGTGGGTCTCCAAGAAAATTACCCCTAAATTC
ACAAAGATTCTACTTTTCTCCACCTCCAATGGATTCCAGAGAGATCCACCACCAACAAC
AGCAACAACAACAACAACAACAGCAGCAGCAGCAACAACAGCAACATCTACAACAACAGC
AACAACCACCGCCAGGGATGTTAATGAGTCACCACAATTCTTACAATCGAAACCTAACG
CCGCCGCCGCTGTTTTAATGGGTCAACACCTCCACATCTCAAGCTATGCATCAAAGAT
TACCTTTTGGTGGTTCTATGTACCGCATCAGCCTCAACAACATCAGTATCATCATCCTC
AGCCTCAGCAACAGATAGATCAGAAGACTCTTGAATCTCTTGGATTCTCTACTTCGCCTC
TTCCTTCTGCTTCTAATTCTTACGGTGGTGGAAATGAAGGAGGTGGTGGTGGTATAGCG
CCGGAGCTAATGCTAACTCTTCCGATCCACCTGCTAAACGGAACAGAGGACGTCCTCTG
GCTCCGGTAAGAAGCAGCTCGATGCTTTAGGAGGAACAGGAGGAGTTGGGTTACGCCCTC
ATGTCATTGAGGTTAAACAGGAGAGGACATAGCTACGAAGATATTGGCGTTTACGAACC
AAGGGCCACGCGCAATCTGTATTCTCTCAGCTACAGGAGCTGTAACTAATGTGATGCTTC
GTCAAGCTAACAATAGCAATCCTACTGGAAGTGTAAAGTATGAGGGCCGATTTGAAATCA
TTTCTCTGTGAGTTCTTTCTTGAATTCTGAGAGTAATGGTACTGTGACCAAACTGGTA
ACTTGAGTGTGTGCTGGCTGGACACGAAGGCCGATTGTGGGTGGATGTGTTGATGGAA
TGCTAGTAGCTGGATCACAAGTCCAGGTCAATTGTGGGAAGCTTTGTACCAGATGGAAGGA
AGCAGAAACAAAGTGCGGGGCGTGCTCAGAATACTCCGGAGCCAGCTTCAGCACCAGCCA
ATATGTTGAGCTTTGGTGGTGTGGTGGACCGGGAAGCCCTCGATCTCAAGGACAACAAC
ACTCGAGCGAGTCATCAGAGGAAAACGAAAGTAATTCTCCGTTGCACCGTAGAAGCAACA
ACAACAACAGCAACAATCATGGGATATTTGGAACTCTACACCTCAACCGCTTCACCAAA
TTCCTATGCAGATGTACCAGAATCTCTGGCCTGGCAACAGTCCTCAATAACAGATGGTT
CATGGGTCAAGATTTGACCGGGTTGCTTCTCTGTTCTTTTGACACATCTCTCCATCAG
ATTTATCTCTATAAAGTAGATTGAGCTCTCTTACTCTCTCATCTTCTCTCTCTTACTAT
TTCTCTTAAATTTAGCTTTGGTTTGTAGATAAATAGAGAGAGAGACATGTTAAGTAGGT
TTCAAATTCATCTTGTGTTTGTCTTAGTAGTTCTTTTGTATTGTGATGATCATA
AAGACTTGTCTTTTCTCTCTATATTCAACGAATTATCCACTTTAA

>G1068 Amino Acid Sequence (domain in AA coordinates: 143-150)

MDSREIHHQQQQQQQQQQQQQQHQQQQPPPGMLMSHHNSYNRNPNAAAVLMGHN
TSTSQAMHQRLPFGGSMSPHQPPQHQQYHHPQPPQQIDQKTLES LGFPTSPLPSASNSYGG

GNEGGGGDSAGANANSSDPPAKRNRGRPPGSGKKQLDALGGTGGVGFTHPHVIEVKTGED
IATKILAFNTNQGPRAICILSATGAVTNVMLRQANNSNPTGTVKYEGRFEIISLSGSFLNS
ESNGTVTKTGNLSVSLAGHEGRIVGGCVDGMLVAGSQVQVIVGSFVPDGRKQKQSAGRAQ
NTPEPASAPANMLSFGGVGGPGSPRSQGGQHSSESSEENESNSPLHRRSNNNNNSNNHGIF
GNSTPQLHLQIPMQMYQNLWPGNSPQ*

>G1225 (1..984)

ATGACTCTAGAAGCTTTATCATCAAACGGTCTTTTAACTTTTGTCTCTGAACTCTT
TCACCAACTCCATTCAAGTCTCTCGTCGATCTCGAGCCATTGCCGGAATGATGTCATC
ATATCGAAGAACAGAATTTTCGGAGATATCTAATCAAGAACCACCACAGCGACAACCA
CCAGCTACGAATCGAGGGAAGAAGCGGCGGAGGAGGAAGCCTAGGGTTTGCAAAAACGAG
GAAGAAGCTGAGAATCAACGAATGACTCACATTGCCGTCGAAAGAAATCGAAGAAGACAA
ATGAATCAACATCTCTCTGTCTTGGCATCTCTCATGCCTCAACCTTTTGTCTACAAGGT
GATCAAGCTTCAATAGTTGGTGGAGCCATAGATTTTCATCAAAGAACTTGAACACAAATTA
CTATCTCTTGAAGCTCAAAAACATCATAATGCTAAATTAACCAGTCGGTTACTTCTTCA
ACAAGTCAAGACTCAATGGTGAACAAGAGAATCCTCATCAACCATCTTCACTATCTCTA
TCGCAGTTCTTTCTTTCATTACATCGATCCGAGCCAAGAGAATAGGAACGGCTCAACAAGC
TCGGTGAACCCCTATGGAAGATCTTGAGGTGACTCTAATCGAACTCATGCTAACATC
AGAATCTTGTCTGAGAGAAGAAGAGGTTTCCGGTGGAGCACGTTGGCCACCACCAACCGCCG
CAGCTTTTGAAGCTGGTGGCTTCTCTACAATCGCTGTCCCTCTCCATTCTTCACCTTAGT
GTCACAACATTGGACAATTATGCTATTTACTCCATCAGCGCTAAGGTGGAAGAGAGTTGC
CAGCTAAGTTTCAGTAGATGACATTGCAGGAGCAGTTCAACCATGCTAAGTATCATTGAA
GAGGAGCCTTTTTGTTGCTCATCAATGTGAGAATTACCATTGACTTCTCTTTGAATCAC
TCAAATGTCACTCATTCTCTCTGAGAAATCTCTTTTTTGTGTTGTTATTCCTTCTTTTA
ATTTTATCACATAGCACATCTTTAGTTTTTTTTTTTT

>G1225 Amino Acid Sequence (domain in AA coordinates: 78-147)

MTLEALSSNGLLNFLSETLSPTPFKSLVDLEPLPENDVLIISKNTISEISNQEPQPQRQP
PATNRGKKRRRRKPRVCKNEEEAENQRMTHIAVERNRRRQMNQHLVLRSLMPQPFQAHKG
DQASIVGGAIDFIKELEHKLSSLEAQKHNAKLNSVTSSTSDSNGEQENPHQPSLSL
SQFFLHSYDPSQENRNGSTSSVKTPMEDLEVTLIETHANIRILSRRRGFRWSTLATTKPP
QLSKLVASLQSLSLILHLSVTLTDNYAIYSISAKVEESCQLSSVDDIAGAVHMLSIIE
EFPCCSSMSELPFDLSLNHNSNVTHSL*

>G1337 (97..1398)

AATGGATTGTGCATCATTCTTCTCACCGTCTCTAGTCTCTGAAAATAAATTCTGATTTTG
ATTTTGAATTTTAGGGATTTTGAGAGAGAGTCAGTTATGAGTAGTTCGGAGAGAGTACCG
TGCGATTTCTGCGGCGAGCGTACGGCGGTTTTGTTTTGTAGAGCCGATACGGCGAAGCTG
TGTTTTGCCCTTGTGATCAGCAAGTTCACACGGCGAATCTGTTGTCGAGGAAGCACGTGCGA
TCTCAGATCTGCGATAATTGCGGTAACGAGCCAGTCTCTGTTCCGGTGTTCACCGATAAT
CTGATTTTGTGTCAGGAGTGTGATTGGGATGTTACCGAAGTGTTCAGTTTCCGATGCT
CATGTTTCGATCCGCCGTGGAAGGTTTTTCCGGTGTCCATCGCGGTTGGAGCTTGCTGCT
TTATGGGGACTTGATTTGGAGCAAGGGAGGAAAGATGAAGAGAATCAAGTTCGATGATG
GCGATGATGATGGATAATTCGGGATGCAAGTTGGATTCTTGGGTTTTGGGATCTAATGAA
TTGATTGTTCCAGCGATACGACGTTTAAAGACGCTGGATCTTGGGATCTAGTTGTGGG
AGGTATAAGCAGGTATTGTGTAAGCAGCTTGAGGAGTTGCTTAAGAGTGGTGTGTCGGT
GGTGTGCGGATGATGGTGTGATCGTGACCGTGATTGTGACCGTGAGGGTGCTTGTGATGGA
GATGGAGATGGAGAAGCAGGAGAGGGGCTTATGGTTCCGGAGATGTCAGAGAGATTGAAA
TGGTCAAGAGATGTTGAGGAGATCAATGGTGGCGGAGGAGGAGGAGTTAACCAGCAGTGG
AATGCTACTACTACTAATCCTAGTGGTGGCCAGAGTTCTCAGATATGGGATTTTAACTTG
GGACAGTCACGGGGACCTGAGGATACGAGTCGAGTGGAAGCTGCATATGTAGGGAAAGGT
GCTGCTTCTTTCATTCAATCAACAATTTTGTGACCATATGAATGAACTTGTTCCTACT
AATGTGAAAGGTGTCAAAGAGATTAAGAAAGGATGACTACAAGCGATCAACTTCAGGCCAG
GTACAACCAACAAATCTGAGAGCAACAATCGTCAATTACCTTTGGCTCTGAGAAAGGT
TCGAACTCCTCCAGTGACTTGCACTAAGGCTGATCTGGAGCGGCTGGCTCAGAACAGAGGAGATGCA
ATGCAGCGTTACAAGGAAAAGAGGAAGACACGGAGATATGATAAGACCATAAGGTATGAA
TCGAGGAAGGCAAGAGCTGACACTAGGTTGCGTGTGAGGAGCAGATTTGTGAAAGCTAGT
GAAGCTCCTTACCCTTAACCTTAAGTTTTTTCACATAGGCTTCTTTTAGCTACAACTT
AGTTACTTTTTTACTCCACTGCCCTCATAAATGTACAGACCGGTCTCGTTTCATCTGGCC

GCCCTTCTTGTTTTATTGCCTTATCTGGCCCTTTTATGTACCTTGAATCTTATCTAGTT
TAAAAAAGATTGTAACCTTCTAGAAAACCATATTCTGTTGACAGTATATACATGTCTATC
CAAGCAAAAA

>G1337 Amino Acid Sequence (domain in AA coordinates: 9-75)
MSSSERVPCDFCGERTAVLFCRADTAKLCLPCDQQVHTANLLSRKHVRSQICDNCNEPV
SVRCFTDNLILCQECDWDVHGSCSVSDAHVRSVEGFSGCPSALELAALWGLDLEQGRKD
EENQVPMAMMDNFMQLDSWVLGSNELIVPSDTTFKKRGSCGSSCGRYKQVLCKQLEE
LLKSGVVGGDGDGDRDRDCDREGACDGDGDGEAGEGLMVPESERLKWSRDVEEINGGG
GGGVNQWNATTNPSSGGQSQIWFNLGQSRGPEDTSRVEAAYVGKGAASSFTINNFDV
HMNETCSTNVKGVKEIKKDDYKRSTSGQVQPTKSESNNRPITFGSEKGSNSSDLHFTEH
IAGTSCKTTRLVATKADLERLAQNRGDAMQRYKEKRKTRRYDKTIRYESRKARADTRLRV
RGRFVKASEAPYP*

>G1759 (110..700)
CGAGAAAAGGAAAAAAAAAATAGAAAGAGAAAACGCTTAGTATCTCCGGCGACTTGAAC
CCAAACCTGAGGATCAAAATTAGGGCACAAAGCCCTCTCGGAGAGAAGCCATGGGAAGAAA
AAAACCTAGAAATCAAGCGAATTGAGAACAAAAGTAGCCGACAAGTCACCTTCTCCAAACG
TCGCAACGGTCTCATCGAGAAAGCTCGTCAGCTTTCTGTTCTCTGTGACGCATCCGTCGC
TCTTCTCGTCTCTCCGCCCTCCGGCAAGCTCTACAGCTTCTCCTCCGGCGATAACCTGGT
CAAGATCCTTGATCGATATGGGAAACAGCATGCTGATGATCTTAAAGCCTTGGATCATCA
GTCAAAGCTCTGAACATATGGTTCACACTATGAGCTACTTGAACCTTGTGGATAGCAAGCT
TGTGGGATCAAATGTCAAAAATGTGAGTATCGATGCTCTTGTTCAACTGGAGGAACACCT
TGAGACTGCCCTCTCCGTGACTAGAGCCAAGAAGACCGAACTCATGTTGAAGCTTGTGA
GAATCTTAAAGAAAAGGAGAAAATGCTGAAAGAAGAGAACCAGGTTTTGGCTAGCCAGAT
GGAGAATAATCATCATGTGGGAGCAGAAGCTGAGATGGAGATGTCACCTGCTGGACAAAT
CTCCGACAATCTTCCGGTGACTCTCCCACTACTTAATTAGCCACCTTAAATCGGCGGTTG
AAATCAAAATCCAAACATATATAATTATGAAGAAAAAAAAAATAAGATATGTAATTATT
CCGCTGATAAGGGCGAGCGTTTTGTATATCTTAATACTCTCTCTTTGGCCAAGAGACTTTG
TGTGTGATACTTAAGTAGACGGAATAAGTCAATACTATCTGTTTTAAGACAAAAGGTTG
ATGAACCTTTGTACCTTATTCGTGTGAGAAAAAAAAAAAAAAAAA
>G1759 Amino Acid Sequence (conserved domain in AA coordinates: 2-57)
MGRKKLEIKRIENKSSRQVTFSKRRNGLIEKARQLSVLCDASVALLVVSASGKLYSFSSG
DNLVKILDYRKQHADDLKDLDHQSALNYGSHYELLELVDSKLVGSNVKNVSDALVQL
EEHLETALSVTRAKKTEMLKLVENLKEKEKMLKEENQVLASQMENNHHVGAEEEMEMSP
AGQISDNLPLVTLPLLN*

>G1804 (169..1497)
TATCTCTCTCTTTCTCAAAACCTTTTCAGTCAAAATTCTCCGGCGGCTTTTAAACTATGTG
AAGGAGGAGAACCTCCATAACAAGAAGCGGATTCTCTCAGTTTTCGGGCGGCGGAGGAAC
ACAAAGCCACCGGTTTTTAGACACACAGATTTCATTTTTCAGTTGTTAAATGGTAACTAGA
GAAACGAAGTTGACGTCAGAGCGAGAAGTAGAGTCGTCATGGCGCAAGCGAGACATAAT
GGAGGAGGTGGTGGTGAGAATCATCCGTTTACTTCTTTGGGAAGACAATCCTCTATCTAC
TCATTGACCTTTGACGAGTTCCAACATGCTTTATGTGAGAACGGCAAGAAGCTTTGGGTCC
ATGAACATGGACGAGTTTCTTGTCTCTATTTGGAACGCAGAGGAGAATAATAACAATCAA
CAACAAGCAGCAGCAGCTGCAGGTTTACATTCTGTTCCGGCTAATCACAATGGTTTCAAC
AACAACAATAACAATGGAGGCGAGGGTGGTGTGGTGTCTTTAGTGGTGGTTCTAGAGGC
AACGAAGATGCTAACAATAAGAGAGGGATAGCGAACGAGTCTAGTCTTCCTCGACAAGGC
TCTTTGACACTTCCAGCTCCGCTTTGTAGGAAGACTGTTGATGAGGTTTGGTCTGAGATA
CATAGAGGTGGTGGTAGCGGTAATGGAGGAGACAGCAATGGACGTAGTAGTAGTAAT
GGACAGAACAATGCTCAGAACGGCGGTGAGACTGCGGCTAGACAACCGACTTTTGGAGAG
ATGACACTTGAGGATTTCTTGGTGAAGGCTGGTGTGGTTAGAGAACATCCCACTAATCCT
AAACCTAATCCAAACCCGAACCAAAACCAAAACCCGCTTAGTGTAATACCCGAGCTGCA
CAGCAACAGCTTTATGGTGTGTTTCAAGGAACCGGTGATCCTTCATTCCCGGGTCAAGCT
ATGGGTGTGGGTGACCCATCAGGTATGCTAAAAGGACAGGAGGAGGAGGGTATCAGCAG
GCGCCACCAAGTTGAGGAGGTTGCTATGGAGGTGGCGTTGGGTTTGGAGCGGGTGGAA
CAGCAAATGGGAATGGTTGGACCGTTAAGCCCGGTGCTTTCAGATGGATTAGGACATGGA
CAAGTGGATAACATAGGAGGTGAGTATGGAGTAGATATGGGAGGGCTAAGGGGAAGGAAA
AGAGTAGTGGATGGTCCAGTGGAGAAAGTAGTGGAGAGAAGACAGAGGAGGATGATCAAG
AACCAGGAGTCTGCTGCTAGATCTAGAGCAAGAAAACAAGCATATACAGTGAATTGGA

GCTGAACCTTAACCAAGTTGAAAGAAGAGAATGCGCAGCTAAAAATGCATTGGCGGAGTTG
GAGAGGAAGAGGAAGCAACAGTATTTTGAAGTCAAGGGCACAACCGAAATTC
CCGAAATCGAACGGGAGATTGCGGACATTGATGAGGAACCCGAGTTGTCCACTCTAAACA
AACAAATAGGAAGATGGAGAAGAAGTCGGAGACAGAACGAGGGAAAACTGATGATTTTCT
ACGTTGTTGTTTTGTCTTTGAGGAATGAGGTTATAGAATCTTTATACTTTGATGTTTTCT
GTGTTGGTAGGAGGAACACCATCTGATCTGCTTTACTAGTGTTCCCTGTGAACAAAGAAA
GTGATTCTGTGTTTCAACATCATCAATCTTTGGAAA

>G1804 Amino Acid Sequence (domain in AA coordinates: 357-407)

MVTRETKLTSEREVESMAQARHNGGGGGENHPFTSLGRQSSIYSLTLDEFQHALCENGK
NFGSMNMDEFLVSIWNAEENNNNQQAAGASHSV PANHNGFNNNNNNGGEGVGVFSG
SSRGNE DANNKRGIANESSLPQGS LTL PAPERKRTVDEVWSEIHRGGSGNGGDSNGRS
SSSNGQNNNAQNGGETAARQPTFGEMTLED FLVKAGVREHPTNPKPNPNPNQNPSSVI
PAAQQQLYGVFQGTGDPSPFGQAMGVGDPSGYAKRTGGGGYQQAPPVQAGVCYGGGVGF
GAGGQQMGMVGPLSPVSSDGLGHGQVDNIGQYGVDMGGLRGRKRVDGPVEKVVERRQR
RMIKNRESAARSARKQAYTVELEAELNQLKEENAQLKHALAELERKRKQYFESLSRA
QPKLPKSNGLRLRTLNRNPSCPL*

>G207 (16..930)

aaaagatctgtttcaatggcggatcgtgttaaaggtccatggagtcagaagaagatgag
cagctacgaaggatggttgagaaatacggaccgaggaattggtctgcgattagcaaatacg
attccaggtcgatctggtaaatcgtgtagattacgttggtgtaatacagttatctccggag
gttgagcatcgtcctttctcgcggaggaagatgagactattgtaaccgcccgtgctcag
tttggttaacaagtggtggcgacgattgctcgtcttcttaacggctcgtagcgataacgccgtt
aaaaatcactggaactctacgcttaagaggaaatgcagcggaggtgtggcggttacgacg
gtgacggagacggaggaagatcaggatcggccgaagaagaggagatctgttagctttgat
cctgcttttgctccggtggatactggattgtacatgagtcctgagagtcctaaccggaatc
gatgttagtgattctagcacgattccgtcaccgtcgtctcctggtgctcagctgtttaaa
ccatgcccgaattccggcggttttacgggtggttccgcagccgttaccggttgaaatgtct
tcgtcttcggaggatccacctaacttcgttgagtttgcactacctggagctgagaacacg
agttcgagccataacaataacaacaacgcgttgatgtttccgagatttgagagtcagatg
aagattaatgtagaggagagaggaggaggaggaaggacgtagaggtgagtttatgacg
gtggtgcaggagatgataaaaagctgaagtgaggagttacatggcgaaatgcagaaaaca
agtgggtggttcgtcgtcggaggtttatacgaatccggcggaatggtggttttagggat
tgtggagtaataacacctaaggttgagtagttttggtttagggttaaaacttgaatcgat
tggggattttcaagagcattcatttttgggggtttatggtaaaattaaaaacaaaaacaaa
atgtacagaggaattaaaatttctatggaataatcttaaatctcaaatatttggtacttg
ttttggtgattcataacccaaatcaaa

>G207 Amino Acid Sequence (domain in AA coordinates: 6-106)

MADRVKGPWSQEEDQLRRMVEKYGPRNWSAISKSI PGRSGKSCRLRWCNQLSPEVEHRP
FSPEEDETIVTARAQFGNKWATIIARLLNGRTDNAVKNHWNSTLKRKCSGGVAVTTVTETE
EDQDRPKKRRSVSFDPAFAPVDITGLYMSPESPNGIDVSDSSTIPSPSSPVAQLFKPMPIS
GGFTVVVPQPLPVEMSSSEDPP TSLSLPGAENTSSSHNNNNNLMFPRFESQM KINVE
ERGGGGEGRRGEFMTTVQEMIKAEVRSYMAEMQKTSGGFVVGGLYESGGNGGFRDCGVIT
PKVE*

>G218 (1..1182)

ATGGAGGCAGAGATCGTGAGACGATCGGAGGTAACGGGATTAAGAAGGGAGGTGGAAGAA
TCGTCAATTGGTAGAGGAGATTGCGATGGTGATGGCGGCGATGTGGGAGAAGATGCGGCA
GGGTTTCGTTGGGACGAGCGGGAGAGGAAGAAGAGATCGAGTTAAAGGGCCGTGGTCGAAG
GAGGAGGATGATGTGTTGAGTGAGCTCGTTAAGAGGTTGGGAGCGAGGAATTGGAGTTTT
ATCGCTCGGAGTATTCCTGGTCGTT CAGGCAAGTCTTGTCGTCTTCGTTGGTGTAATCAG
CTCAATCCAAATCTTATACGCAATTCATTTACTGAGGTAGAGGATCAGGCTATCATCGCA
GCACATGCCATCCACGGAACAAATGGGCTGTTATCGCGAAGCTCCTCCCGGAAGAACA
GATAATGCTATCAAGAACCACTGGAACCTCTGCTTTAAGACGTCGATTCATAGACTTTGAA
AAGGCCAAGAAATAGGAACTGGAAGCTTGGTCGTGGATGATCTGGATTTGACAGAACG
ACAACAGTAGCCTCATCAGAAGAACTTTATCTTCAGGCGGTGGTTGCCATGTAAC TACT
CCAATTGTATCTCCAGAAGGCAAAGAAGCTACCACCTCCATGGAAATGTCTGAAGAACAA
TGCGTAGAGAAAACAAACGGAGAAGGTATTTCTAGGCAAGATGATAAGGATCCTCCACG
CTTTTCCGCCCAAGTGCCTCGGCTCAGTTCTTTAATGCTTGCAATCACATGGAAGGATCA

CCCTCTCCACATATACAAGACCAAAATCAGCTCCAATCATCTAAACAAGACGCAGCAATG
CTAAGATTGCTTGAAGGAGCTTACAGCGAACGGTTTGTGCCTCAAACATGTGGAGGTGGT
TGTTGCAGCAACAATCCCGATGGCAGTTCCTCAGCAAGAATCATTGTTGGGTCCAGAGTTT
GTGGATTACTTAGACTCACCAACGTTTCCGAGTTCGGAAGTCTGCTATAGCAACGGAA
ATAGGCAGCCTCGCTTGGCTGAGAAGCGGTTTAGAGAGTAGCAGCGTGAGGGTGATGGAA
GACGCAGTTGGTCGGTTAAGGCCTCAAGGCTCCAGGGGTATCGAGATCATTATCTTGTA
TCTGAACAGGGGACGAACATAACCAATGTCCTGTCCACATAA

>G218 Amino Acid Sequence (domain in AA coordinates: TBD)

MEAEIVRRSEVTGLRREVEESSIGRGDCDGDGVDGEDAAGFVGTSGRGRDRVKGPWSK
EEDDLVSELVKRLGARNWSFIARSIPGRSGKSCRLRWCNQLNPNLIRNSFTEVEDQAIIA
AHAIHGNKWAIVIAKLLPGRTDNAIKNHWSALRRRFIDFEKAKNIGTGLVVDSDGFDRT
TTVASSEETLSSGGGCHVTTPIVSPEGKEATSMEMSEEQCVEKTNGEGISRQDDKDPPT
LFRPVPRLSSFNACNHMEGSPSPHIQDQNLQSSKQDAAMLRLLEGAYSERFVPQTCCGG
CCSNPDGSGFQESLLGPEFVDYLDSPFPSSELAIAIEIGSLAWLRSGLESSSVRVME
DAVGRLRPQSGRGRDHYLVSEQGTNITNVLST*

>G241 (46..867)

GAAAAACATTTCACTTCTTTTATCAGCAATCACAATCAAAGAGATGGGAAGAGCTCCA
TGCTGTGAGAAGATGGGGTTGAAGAGAGGACCATGGACACCTGAAGAAGATCAAATCTTG
GTCTCTTTTATCCTCAACCATGGACATAGTAAGTGGCGAGCCCTCCCTAAGCAAGCTGGT
CTTTTGAGATGTGGAAGAGCTGTAGACTTAGGTGGATGAAGTATTTAAAGCCTGATATT
AAACGTGGCAATTTACCAAAGAAGAGGAAGATGCTATCATCAGCTTACACCAAATACTT
GGCAATAGATGGTCAGCGATTGCAGCAAACTGCCTGGAAGAACCATAACGAGATCAAG
AACGTATGGCACACTCACTTGAAGAAGAGACTCGAAGATTATCAACAGCTAAACCTAAG
ACCAGCAACAAAAAGAGGGTACTAAACCAAATCTGAATCCGTAAATAACGAGCTCGAAC
AGTACTAGAAGCGAATCGGAGCTAGCAGATTTCATCAAACCTTCTGGAGAAAGCTTATTT
TCGACATCGCCTTCGACAAGTGAGGTTTCTTCGATGACACTCATAAGCCACGACGGCTAT
AGCAACGAGATTAATATGGATAACAAACCGGGAGATATCAGTACTATCGATCAAGAATGT
GTTTCTTTTCGAACTTTTGGTGCGGATATCGATGAAAGCTTCTGGAAAGAGACTGTAT
AGCCAAGATGAACACAACTACGTATCGAATGACCTAGAAGTCGCTGGTTTAGTTGAGATA
CAACAAGAGTTTCAAACTTGGGCTCCGCTAATAATGAGATGATTTTTCAGTGAGATG
GAAGTCTGGTTCGATGTATTGGCTAGAACCGGCGGGGAACAAGATCTCTAGCCGGGCT
CTAGTTAACATGTTTGAGGAGTAAAGTGAAATGGTGCAAATTAGTTAAGGCTAAGAAATT
CAAAAGCTTTTGTGTTTACCGAGAAAAAACACACTCTAACTCTTGATGTGATGTAGTTAGT
GTATTAATTAGAGGCTGCGTTTTCAA

>G241 Amino Acid Sequence (domain in AA coordinates: 14-114)

MGRAPCCEKMLKRPWPTEEDQILVSFILNHGHSNWRALPKQAGLLRCGKSCRLRWMNY
LKPDIKRGNFTKEEEDAIISLHQILGNRWSAIAKLPGRDTDNEIKNVVHHLKRLLEDYQ
PAKPKTSNKKKGTKPKSESVITSSNSTRSESELADSSNPSGESLFTSPSTSEVSSMTLI
SHDGYSNEINMDNKPGLDISTIDQECVSFETFGADIDESFWKETLYSQDEHNYVSNLDLEVA
GLVEIQQEFQNLGSANNEMIFDSEMELLVRCIG*

>G254 (15..923)

CGATTTTCGAGCTCTATGGTGTCCGTAAACCTAGACCTAAGGGTTTTCCAGTTTTCGATT
CCTCGAATATGAGTTTACCAAGCTCCGATGGATTGGTTTCGATTCCGGCCACGGGACGGA
CCAGTACGGTGTCTGTTTTCTGAGGATCCGACGACGAAGATTCCGAAGCCGTACACAATCA
AGAAGTCGAGAGAGAATTGGACAGATCAAGAGCACGATAAATTTCTAGAAGCTCTTCACT
TATTCGATAGGGATTGGAAGAAAAATAGAAGCCTTTGTTGGATCAAAAACAGTAGTTTCA
TACGAAGCCACGCTCAGAAATACTTTCTCAAAGTTTCAAGAGTGGTGTACCAACATC
TTCCACTTCTCGACCTAAGAGGAAAGCGAGTCATCCTTATCCTATAAAGGCTCCTAAAA
ATGTTGCTTATACCTCTCTCCGCTCTTCGAGTACATTACCGTTGCTTGAGCTGGTTATT
TGATATAGCTCTGATTTCGAAGTCATTGATGGGAAACCAAGCTGTTTGTGCATCTACCTCTT
CTTCGTGGAATCATGAATCGACAAATCTGCCAAAACCGGTGATTGAAGAGGAACGGGAG
TCTCGGCCACGGCTCTCTCCCAAATAATCGCTGCAGACAGGAAGATACAGAGAGGGTAC
GAGCAGTGACAAAGCCAAATAACGAAGAAAGTTGTGAAAGCCACATAGAGTGATGCCGA
ATTTTGCTGAAGTTTACAGCTTCATTGGAAGTGTCTTCGATCCCAACACATCAGGCCACC
TCCAGAGATTAAAGCAGATGGATCCAATAAATATGGAAACGGTCTTTTACTGATGCAAA
ACCTGTCTGTAAATCTGACAAGTCCCGAGTTTGACAGACAAAGGAGGTTGATATCATCAT
ACAGCGCTAAAGCTTTGAAATAGAGATAGAATAAAACAATAATGTACCTTATGTGAGATC

AAGAGACAATCATCCAAGGTCTGTATGCATTGCTTGGATTTAGGCCTCGTGTTCTCACTA
CAGGAGCAGAACCAATCGCAAAGACTCTTAGATGGCTACTGAGTTGTGGTTTTTATGTCT
CTGTAAGTCGCGGTGGAGCACACGTGTTTGTCTGTCTTGTGTATGTGTGTATAGATAAT
ACAAGGTTTTGTCAGAGTAAGGTACAGTTAGCTGCAAGTGAGTTTGGATCAATCTTAAGA
TTAAAACCTGAGAGTGAGTGTCCAAAGAGACTGTGTAATATTGGTTTGGCGGTGAGCAG
AAGAGTTTTGAAGTGCACATCCAGTTAGTGATAACACGGTTGAAGAAAAGGTAAGGTTAC
AAGTTTAGTTTTGAATAATTGTATACTCAAAAAATATGAATGTATAAAGAATAATCACTT
GAGTCGCCTTA

>G254 Amino Acid Sequence (domain in AA coordinates: 62-106)
MVSVNPRPKGFVFDSSNMSLPSSDGFSGIPATGRTSTVSFSEDPTTKIRKPYTIKKSRE
NWTQEHDKFLEALHLFDRDWKKIEAFVGSKTVVQIRSHAQKYFLKVQKSGANEHLPLPR
PKRKASHPYPIKAPKNVAYTSLPSSSTLPLLEPGYLYSSDSKSLMGNQAVCASTSSSWNH
ESTNLPKPVIEEPGVSATAPLPNNRCRQEDTERVRAVTKPNNEESCEKPHRVMPNFAEV
YSFIGSVFDPNTSGHLQRLKQMDPINMETVLLLMQNLSVNLTSPEFAEQRLISSYSAKA
LK*

>G26 (73..729)
TTGGCTTGTAACCAACCCATCTTTGACTTCAAAAATAAAAATAAATCATAATTGA
CATCATCGGATAATGCATAGCGGGAAGAGACCTCTATCACCAGAATCAATGGCCGGAAAT
AGAGAAGAGAAAAAAGAGTTGTGTTGTTGCTCAACTTTGTGCGGAATCTGATGTGTCTGAT
TTTGTCTCTGAACCTACTGGTCAACCCATCCCATCATCCATTGATGATCAATCTTCGTCG
CTTACTCTTCAAGAAAAAAGTAACCTCGAGGCAACGAACTACAGAGGCGTGAGGCAAAGA
CCGTGGGGAAAAATGGGCGGCTGAGATTTCGTGACCCGAACAAGGCAGCTCGTGTGTGGCTT
GGGACGTTTCGACACTGCAGAAGAAGCCGCCCTTAGCGTATGATAAAGCTGCATTTGAGTTT
AGAGGTCAACAAGGCCAAGCTTAACCTCCCGGAGCATATTCGTGTCAACCCCTACTCAACTC
TATCCATCGCCCGCTACTTCCCATGATCGCATTATCGTGACACCACCTAGTCCACCTCCA
CCAATTGTCTCCTGACATACTTCTTGATCAATATGGCCACTTTCAATCTCGAAGTAGTGAT
TCCAGTGCCCAACTTGTCCATGAATATGCTGTCTTCTTCGTCTTCATCTTTGAATCATCAA
GGGCTAAGACCAAATTTGGAGGATGGTGAAAACGTGAAGAACATTAGTATCCACAAACGA
CGAAAATAACATGTTAATGGCATAAATATCTCTTCGTCCAAGTTATCAAACGCATTGACC
TCCGGCTTTGATCATTTTTAGGCGCTTAATCTCTTTACGACTTCATTTTGGTAGTCTTTAA
AGAGTCTATGGAGTGGATTTAGCTAGGAATCAGGCCTTATGGATGAAAAATATATAAATT
TTGAACATGACTATGCAAGAATGGGATGAAGACTACTTAGCTTGAAAAACGTCCTGATAG
GTCATGACGACTATATCCACAGAAGATGACCGACGGAGACAACAACATGCCTCACCTGAT
CGACCGATCAAATGAGATAATGTGTTGACCGGACCGGTCCGATCAGGTTGGGTGAGTAT
ATCA

>G26 Amino Acid Sequence (domain in AA coordinates: 67-134)
MHSGKRPLSPESMAGNREEKELCCCSTLSESDVSDVFVSELTGQPIPSIDQSSSLTLQ
EKSNSRQRYRQVRQRPWGKWAABIRDPNKAARVWLGTDFDTEEAALAYDKAAFEFRGHK
AKLNFPEHIRVNPTQLYSPATSHDRIIVTPSPPPPIAPDILLDQYGHFQSRSSDSSAN
LSMNMSSSSSSSLNHQGLRPNLEDEGENVKNISIHKRRK*

>G263 (48..902)
TTTTTAGTTTATTTTTCTGTGGTAAAAATAAAAAAGTTCCGCCGAGATGACGGCTGTGA
CGGCGGCGCAAAGATCAGTTCCGGCGCCGTTTTTAAGCAAACGTATCAGCTAGTTGATG
ATCATAGCACAGACGACGTCGTTTCATGGAACGAAGAAGGAACAGCTTTTGTCTGTGGA
AAACAGCAGAGTTTGCTAAAGATCTTCTTCTCAATACTTCAAGCATAATAATTCTCAA
GCTTCATTTCGTGAGCTCAACACTTACGGATTTTCGTAAACTGTACCGGATAAATGGGAAT
TTGCAAACGATTATTTCCGGAGAGGCGGGGAGGATCTGTTGACGGACATACGACGGCGTA
AATCGGTGATTGCTTCAACGGCGGGGAAATGTGTTGTGTTGGTTTCGCCTTCTGAGTCTA
ATTCTGGTGGTGGTGATGATCACGGTTCAAGCTCCACGTCATCACCCGGTTTCGTGGAAGA
ATCCTGGTTTCGGTGGAGAACATGGTTGCTGATTATCAGGAGAGAACGAGAAGCTTAAAC
GTGAAAACAATAACTTTGAGCTCGGAGCTCGCGGCGGCGAAGAAGCAGCGCATGAGCTAG
TGACGTTCTTGACGGGTCATCTGAAAGTAAGACCGGAACAAATCGATAAAATGATCAAAG
GAGGGAAATTTAAACCGGTGGAGTCTGACGAAGAGAGTGAGTGCGAAGGTTGCGACGGCG
GCGGAGGAGCAGAGGAGGGGTAGGTGAAGGATTGAAATTGTTTGGGGTGTGGTTGAAAG
GAGAGAGAAAAAGAGGGACCGGGATGAAAAGAATTATGTGGTGGTGGTCCCGTATGA
CGGAAATAAAGACGTGGACTTTACGCGCCGTTGTGGAAGAGCAGCAAAGTCTGCAACT
AAAAAAGAGTAGAAGACTGTTCAAACAGCGTGTGACACGTCATCGACGACGAGAAAA

AAATGATTTAAAAAACTATTTTTTCCGTAAGGAAGAAAAGTTATTTTTATGTTTTAAAA
AGGTGAAGAAGGTCCAGAAGGATCAACGCAATATATAAATGGATTTTCATGTATTATAT
AATTTAATTAGTGTATTAAGAAAA

>G263 Amino Acid Sequence (domain in AA coordinates: TBD)

MTAVTAAQRSVPAPFLSKTYQLVDDHSTDDVVSWNEEGTAFVWVKTAFAKDLLPQYFKH
NNFSSFIQRLNTYGRKTVDPDKWEFANDYFRGGEDLLDIRRRKSVIASTAGKCVVGS
PSESNSGGDDHGSSTSSPGSSKNPGSVENMVADLSGENEKLKRENNNLSSSELAACKQ
RDELVTFLTGHKLVRPEQIDKMIKGGKFKPVESDEESECEGCDGGGGAEEGVGEGLKLFG
VWLKGERKKRDRDEKNYVVSGRMTEIKNVDFHAPLWKSSKVCN*

>G308 (196..1794)

AGTAATTTAGTTTTTTTTTTTTTTTTTACAAATTTATTTTGTATTAGAAAGTGGTAGTGG
AGTGAAAAAACAAATCCCTAAGCAGTCCCTAACCGATCCCCGAAGCTAAAGATTCTTCACCT
TCCCAAATAAAGCAAAACCTAGATCCGACATTGAAGGAAAAACCTTTTAGATCCATCTCT
GAAAAAAACCCCAACCATGAAGAGAGATCATCATCATCATCAAGATAAGAAGACTATG
ATGATGAATGAAGAAGACGACGGTAACGGCATGGATGAGCTTCTAGCTGTTCTTGTTAC
AAGGTTAGGTCATCGGAAATGGCTGATGTTGCTCAGAAACTCGAGCAGCTTGAAGTTATG
ATGTCTAATGTTCAAGAAGACGATCTTCTCAACTCGCTACTGAGACTGTTCACTATAAT
CCGGCGGAGCTTTACACGTGGCTTGATTCTATGCTACCGACCTTAATCCTCCGTCGTCT
AACGCCGAGTACGATCTTAAAGCTATTTCCCGGTGACGCGATTCTCAATCAGTTCGCTATC
GATTCGGCTTCTTCTCTAACCAAGGCGGCGAGGAGATACGTATACTACAAACAAGCGG
TTGAAATGCTCAAACGCGCTCGTGGAACACCACAGCGACGCGCTGAGTCAACTCGGCAT
GTTGTCCTGGTTGACTCGCAGGAGAACGGTGTGCGTCTCGTTCACGCGCTTTTGGCTTGC
GCTGAAGCTGTTTCAAGGAGAATCTGACTGTGGCGGAAGCTCTGGTGAAGCAAATCGGA
TTCTTAGCTGTTTCTCAAATCGGAGCTATGAGACAAGTCGCTACTTACTTCGCCGAAGCT
CTCGCGCGGCGGATTACCGTCTCTCTCCGTCGCAGAGTCCAATCGACCACTCTCTCTCC
GATACTCTTCAGATGCACCTTCTACGAGACTTGTCTTATCTCAAGTTCGCTCACTTCACG
GCGAATCAAGCGATTCTCGAAGCTTTTCAAGGGAAGAAAAGAGTTCATGTCAATTGATTTC
TCTATGAGTCAAGGTCTTCAATGGCCGCGCTTATGCAGGCTCTGCGCTTCGACCTGGT
GGTCTCTCTGTTTTCCGGTTAACCAGGAATTGCTCCACCGGCACCGGATAATTTGATTAT
CTTCATGAAGTTGGGTGTAAGCTGGCTCATTTAGCTGAGGCGATTACCGTTGAGTTTGAG
TACAGAGGATTTGTGGCTAACACTTTAGCTGATCTTGATGCTTCGATGCTTGAGCTTAGA
CCAAGTGAGATTGAATCTGTTGCGGTTAACTCTGTTTTCGAGCTTCACAAGCTCTTGGA
CGACCTGGTGCGATCGATAAGGTTCTTGTTGTTGTTGATTAAGGAGATTAAACCGGAGATTTTC
ACTGTGGTTGAGCAGGAATCGAACATAATAGTCCGATTTTCTTAGATCGGTTTACTGAG
TCGTTGCATTATTACTCGACGTTGTTTGAAGTGTACCGAGTGGTCAAGAC
AAGGTCATGTCGGAGTTTACTTGGGTAAACAGATCTGCAACGTTGTGGCTTGTGATGGA
CCTGACCGAGTTGAGCGTCATGAAACGTTGAGTCAGTGGAGGAACCGGTTCCGGTCTGCT
GGGTTTGGCGCTGCACATATTGGTTCAATGCGTTTAAGCAAGCGAGTATGCTTTTGGCT
CTGTTCAACGGCGGTGAGGGTTATCGGGTGGAGGAGAGTGACGGCTGTCTCATGTTGGGT
TGGCACACACGACCGCTCATAGCCACCTCGGCTTGGAACTCTCCACCAATTAGATGGTG
GCTCAATGAATTGATCTGTTGAACCGTTATGATGATAGATTTCCGACCGAAGCCAACT
AAATCCTACTGTTTTTCCCTTTGTCACTTGTTAAGATCTTATCTTTCATTATATTAGGTA
ATTGAAAAATTTTAATCTCGCCTAAATTACT

>G308 Amino Acid Sequence (domain in AA coordinates: 270-274)

MKRDHHHHQDKKTMMEEDDGNGMDELLAVLGKVRSEMAADVAQKLEQLEVMMSNVQ
EDDLSQLATETVHYNPAELYTWLDSMLTDLNPPSSNAEYDLKAIPGDAILNQFAIDSASS
SNQGGGGDTYTTNKRKCSNGVVETTTATAESTRHVVVLVDSQENGVRVLVHALLACAEAVQ
KENLTVAEALVKQIGFLAVSQIGAMRQVATYFAEALARRIYRLSPSQSPIDHSLSDTLQM
HPLYETCPYLKFAHFTANQALILEAFQKKRVHVIDFSMSQGLQWPALMQALALRPGGPPVF
RLTGIGPPAPDNFDYIHEVGKLAHLAEAIHVEFEYRGFVANTLADLDASMLELRPSEIE
SVAVNSVFEHLKLLGRPGAIDKVLGVVNQIKPEIFTVVEQESNHNSPIFLDRFTESLHY
STLFDSLEGVPSGQDKVMSEVYLKGQICNVVACDGPDRVERHETLSQWRNRFSGAGFAAA
HIGSNAFKQASMLLALFNGGEGYRVEESDGCLMLGWHTRPLIATSARKLSTN*

>G38 (149..1156)

GAGGAAACTCGAAAAAGCTACACACAAGAAGAAGAAAAGATACGAGCAAGAAGACT
AAACACGAAAGCGATTATCAACTCGAAGGAAGAGACTTTGATTTCAAATTTCTGTCCTCC
TATAGATTGTGTTTCTGGGAAGGAGATGGCAGTTTATGATCAGAGTGGAGATAGAAA

CAGAACACAAATTGATACATCGAGGAAAAGGAAATCTAGAAGTAGAGGTGACGGTACTAC
TGTGGCTGAGAGATTAAAGAGATGGAAAGAGTATAACGAGACCGTAGAAGAAGTTTCTAC
CAAGAAGAGGAAAGTACCTGCCGAAAGGGTCGAAGAAGGGTTGTATGAAAGGTAAAGGAGG
ACCAGAGAATAGCCGATGTAGTTTCAGAGGAGTTAGGCAAAGGATTTGGGGTAAATGGGT
TGCTGAGATCAGAGAGCCTAATCGAGGTAGCAGGCTTTGGCTTGGTACTTTCCCTACTGC
TCAAGAAGCTGCTTCTGCTTATGATGAGGCTGCTAAAGCTATGTATGGTCCTTTGGCTCG
TCTTAATTTCCCTCGGTCTGATGCGTCTGAGGTTACGAGTACCTCAAGTCAGTCTGAGGT
GTGTACTGTTGAGACTCCTGGTTGTGTTCATGTGAAAACAGAGGATCCAGATTGTGAATC
TAAACCCCTTCTCCGGTGGAGTGGAGCCGATGTATTGTCTGGAGAATGGTGCAGGAAGAGAT
GAAGAGAGGTGTTAAAGCGGATAAGCATTGGCTGAGCGAGTTTGAACATAACTATTGGAG
TGATATTCTGAAAGAGAAAGAGAAACAGAAGGAGCAAGGGATTGTAGAAAACCTGTCAGCA
ACAACAGCAGGATTTCGTATCTGTTGCAGACTATGGTTGGCCCAATGATGTGGATCAGAG
TCACTTGGATTCTTCAGACATGTTTGATGTCTGATGAGCTTCTACGTGACCTAAATGGCGA
CGATGTGTTTGCAGGCTTAAATCAGGACCGGTACCCGGGGAACAGTGTGGCCAAACGGTTC
ATACAGGCCCGAGAGTCAACAAAGTGGTTTTGTATCCGCTACAAAGCCTCAACTACGGAAT
ACCTCCGTTTCAGCTCGAGGGAAAGGATGGTAATGGATTCTTCGACGACTTGAGTTACTT
GGATCTGGAGAACTAAACAAAACAATATGAAGCTTTTGGATTGTATTTGCCTTAATC
CCACAACGACTGTTGATTCTCTATCCGAGTTTTAGTGATATAGAGAACTACAGAACACGT
TTTTTCTTGTATAAAGGTGAAGTGTATATATCGAAACAGTGATATGACAATAGAGAAGA
CAACTATAGTTTGTAGTCTGCTTCTCTTAAGTTGTTCTTTAGATATGTTTTATGTTTTG
TAACAACAGGAATGAATAATACACACTTGTGAAGCTTTAAAAAAAAAAAAAAAAAAAAA
>G38 Amino Acid Sequence (domain in AA coordinates: 76-143)
MAVYDQSGDRNRTQIDTSRKRKSRSGDGTVAERLRKRWKEYNETVEEVSTKKRKVPAKG
SKKKGCMKKGKGPENSRCSFRGVRQRIWGWVAEIREPNRGSRLWLGTFTPAQEAAASAYDE
AAKAMYGPLARLNFPRSDASEVTSTSSQSEVCTVETPGCVHVKTEDPDCEKPFSGGVEP
MYCLENGAEEMKRGVKADKHWLSEFEHNYWSDILKEKEKQKEQIVETCQQQQQDLSVA
DYGWPNDVDQSHLDSSDMFVDVDELLRDLNGDDVFAGLNQDRYPGNSVANGSYRPESQQSG
FDPLQSLNYGIPPFQLEGKDGNGFFDDLSDLDLEN*

>G43 (38..643)

CTCCTGTCTTGTCTAAAGAAAAAGAGAGAGGAAGAAATGGAGACTTTTGAGGAAAGCTC
TGATTTGGATGTTATACAGAAACATCTATTGAAGACTTGATGATCCCTGATGGTTTCAT
TGAAGATTTTGTCTTTGATGATACTGCTTTTGTCTCCGACTCTGGTCTCTAGAACCCTT
TAACCCAGTTCGAAACTGGAACCTAGTTCACCTGTTCTTGATCCAGATTCTATGTCCA
AGAGATTCTGCAAAATGGAAGCAGAATCATCATCATCATCAACAACAACGTACCTGA
GGTTGAGACTGTCTCAAACCGGAAAAAACAAGAGGTTTGAAGAAACGAGACATTACAG
AGGCGTGAGAAGGAGGCCATGGGGGAAATTTGCAGCAGAGATTTCGAGATCCGGCAAAGAA
AGGATCCAGGATTTGGTTAGGCATTTTGTAGAGTGATATTGATGCTGCAAGGGCTTACGA
CTATGCAGCTTTTAAAGCTCAGGGGAAGAAAGCTGTTCTCAACTTTCCTTTGGATGCCGG
AAAGTATGATGCTCCGGTCAATTCATGCCGAAAAAGGAGGAGAACCGATGTACCACAGCC
TCAAGGAACAACAAGTACTTCATCATCGTCATCAAACTAATGGGGGAATAGTGATGT
TTAATTAGTATATATAGGTTAATATCTTAAGTATGTGAAGCATCATGTATAGAGCCAAGA
ACCTGTTAGACTAGTGTACTGAAAAGAACTCTTGCAAAATATGTACTAAAGAGTTCCTGT
AACAATGGAACCTTCTGCGTTTTCTCTTGTCTTAAAGAGCTTAAGGTTCTAGAAACAAAGT
TCTTGTCTTTTCGGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAA

>G43 Amino Acid Sequence (domain in AA coordinates: 104-172)
METFEESDLVDVIQKHLFEDLMIPDGFIEDFVDDTAFVSGLWSLEPFNPVPKLEPSSPV
LDPDSYVQEIQLMEAESSSSSTTSPEVETVSNRKKTKRFEETRHYRGVRRRPWGKFAA
EIRDPAKGSRILWLTFFESDIDAARAYDYAAFKLGRKAVLNFLDAGKYDAPVNSCRKR
RRTDVPQPGTTTSTSSSSN*

>G536 (1..768)

ATGTCGACAAGGGAAGAGAATGTTTACATGGCGAAATTAGCCGAACAAGCTGAACGTTAC
GAAGAAATGGTTGAATTCATGGAGAAAGTTGCGAAAACCTGTTGATGTTGAGGAACCTTCA
GTTGAAGAGAGGAATCTTCTCTCTGTTGCTTACAAGAACGTGATTGGAGCGAGAAGAGCT
TCGTGGAGAATCATTTCTTCGATTGAGCAGAAAGAAGAGAGCAAAGGGAACGAAGATCAT
GTTGCTATTATCAAGGATTACAGAGGAGAGATTGAATCCGAGCTTAGCAAAATCTGTGAT
GGGATTTTGAATGTTCTTGAAGCTCATCTTATTCCTTCTGCTTCACCAGCTGAATCTAAA

GTGTTTTATCTTAAGATGAAGGGTGATTATCATAGGTATCTTGCTGAGTTTAAGGCTGGT
GCTGAAAGGAAAGAAGCTGCTGAAAGCACTTTGGTTGCTTACAAGTCTGCTTCCGACATT
GCCACTGCTGAGTTAGCTCCTACTCACCCGATAAGGCTTGGTCTTGCACTCAACTTCTCT
GTGTTTTACTATGAAATCCTCAACTCGCTGATCGTGCTTGCAAGCTCGCAAAGCAGGCG
TTTGATGATGCAATCGCTGAGTTAGATACATTGGGTGAGGAATCATACAAGGACAGTACA
CTGATTATGCAGCTTCTTAGAGACAATCTCACTCTCTGGACTTCAGATATGACTGACGAA
GCAGGAGATGAGATTAAAGGAGGCATCAAAGCCCGATGGTGCCGAGTAA

>G536 Amino Acid Sequence (domain in AA coordinates:226-233)
MSTREENVYMAKLAEQAERYEEMVEFMEKVAKTV DVEELSVEERNLLSVAYKNVIGARRA
SWRIISSIEQKEESKGNEDHVAIIKDYRGEIESELSKICDGLNLVLEAHLIPSASPESK
VFYLMKMGDYHRYLAEFKAGAERKEAAESTLVAYKSASDIATAELAPTHPIRLGLALNFS
VFYYEILNSPDRACSLAKQAFDDAIAELDTLGEESYKDS TLIMQLLRDNLTLWTSMDTDE
AGDEIKEASKPDGAE*

>G567 (38..1273)
AAAAAGAAGAAATCAGAAAGTGA AAAAGAGAGCGAGCGATGAACAGTATCTTCTCCATTGA
CGATTTCTCCGATCCTTTCTGGGAAACTCCTCCGATTCTCTCAATCCCGACTCTTCTAA
GCCTGTACGGCGGATGAAGTTAGCCAGAGTCAACCGGAATGGACTTTCGAGATGTTTCT
CGAAGAGATTTCTTCGTGCGGCGGTGAGCTCTGAGCCACTTGTTAAACAACAACGCGAT
CGTCGGTGTCTTCTCGGCGCAATCTCTTCTCTGTTTCCGGACAGAATGATTTGAGGA
TGATAGTCGATTTCTGATCGCGATTCCGGAAATTTGGATTGTGCTGCTCCCATGACGAC
GAAGACGGTGAATGTTGATTCCGATGATTATCGTCGTGTTCTTAAGAACAAGCTTGAGGC
TGAGTGCGCGACTGGTGTCTCTTCTCGGGTGGGTCTGTGAAGCCTGAAGATTGACTAG
TTCTCCAGAACTCAACTTCAACAGTTCAATCCAGTCCCTTACTCAAGGAGAAGTTGG
TGTTACTTCTTCTTACCAGCTGAGGTGAAAAAACTGGTGTATCAATGAAGCAGGTTAC
TAGTGGATCGTCGAGAGAAATTTCTGATGACGAGGACCTTGATGAAGAGAATGAACCAC
CGGTTCTTGAAGCCAGAGGACGTTAAAAAATCTAGAAGGATGCTGTCAAATCGTGAGTC
AGCTAGGCGATCTAGAAGGAGAAAGCAGGAGCAACAAGTGACCTCGAAACACAGGTTAA
TGATCTAAAAGGTGAGCATTATCACTTCTTAAACAAGTGAAGCAATGAATCACAAGTA
TGACGAGGCTGCTGTTGGCAATAGAACTAAAGGCTGACATTGAGACATTAAGAGCTAA
GGTGAAAAATGGCGGAAGAAACCGTGAAGAGAGTAACAGGAATGAATCCGATGCTTCTCGG
AAGATCAAGTGGACATAACAACAACAAGAATGCCAATAACTGGTAACAACAGGATGGA
TTCTTCTAGCATTATTCCAGCTTATCAACCACACTCAAACCTAAACCATATGTCAAACCA
AAACATCGGGATCCCAACATTCTACCTCCAAGACTCGGAAACAATTTGCTGCTCCTCC
ATCCCAAACCAAGCTCTCCCTTGCGAGAAATTAAGAAATGGGCAAAATCACCATGTTACTCC
AAGCGCCAACCCGTATGGCTGGAATACCGAACCTCAGAACCGATTGAGCATGGCCGAAAAA
ATGCGTGGACTGATCAACAAGAAGCGGGTTTCGCACTATATTAATGTCTATGCATCTGT
AATTTGTAAGTGTTATTAAGTTACGAATCATGAGAAAACATCTTGTGAAAATACAGTCTC
ATGGCTTATATATATATATAAGCTCTGTCTTATAACATTACAAGATTCTTATTTGAGAA
CGTCTTCTATTTATAGCTAATAAAAAAAAAAAAAAAAAA

>G567 Amino Acid Sequence (domain in AA coordinates 210-270)
MNSIFSIDFSDPFWETPPIPLNPDSSKPVTADEVQSQPEWTFEMFLEEISSAVSSEP
LGNNNNNAIVGVSSAQSLPSVSGQND FEDDSRFRDRDSGNLDCAPMTTKTVNVDSDDYRR
VLKNKLEAECATGVSLRVGSVKPEDSTSSPETQLQPVQSSPLTQGELGVTSSLPAEVKKT
GVSMKQVTS GSSREYSDDEDLDEENETGSLKPEDVKKSRRMLSNRESARRSRRRKQEQ
SDLETQVNDLKGEHSSLLKQLSNMNHKYDEAAVGNRI LKADIETLRKVMAEETVKRVT
GMNPMLLGRSSGHNNNNRMPITGNRMDSSSIIPAYQPHSNLNMNSNQNIPTILPPRL
GNNFAAPPSTSSPLQIRINGQNHVTPSANPYGWNTPEQNDSAWPKKCDV*

>G680 (338..2275)
CAGTTATCTTCTTCTCTCTCTGTTTTTAAATTTATTTT TAGAGAATTTTTTTT
TTTTGCTTCCGATTTGATTATTTCCGGGAACGATGACTTCTCCGGGGAGTTCCCGGTGAG
ATGATAAGTCAGATTGCATATTGTCTCTCCATGGCTACTCTCAAGGGTTTTGGCTGCG
GTGGATTGCTTTGGTTTCTCTAGAATCTAAAGAGGTTATCACAACGGCTTTGCAATTTGA
AACTTTTCATGTTTGGGGAGATCAAAGATGGTTTCTTTTTTATACTTTACTTGTAGAGA
GGATTGGAAGCAGCGAATAGCTGCAACCGGTCTGTTATGGATACTAATACATCTGGAGA
AGAAATATTAGCTAAGGCAAGAAAGCCATATACAATAACAAAGCAGCGAGAGCGATGGAC
TGAGGATGAGCATGAGAGGTTTCTAGAAGCCTTGAGGCTTTATGGAAGAGCTTGGCAACG
AATTGAAGAACATATTGGGACAAAGACTGCTGTTT CAGATCAGAAGTCATGCACAAAAGTT

CTTCACAAAGTTGGAGAAAAGAGGCTGAAGTTAAAGGCATCCCTGTTTGCCAAGCTTTGGA
CATAGAAATTCCGCCCTCCTCGTCTAAACGAAAAACCACTCCTTATCCTCGAAAAACC
TGGGAACAACGGTACATCTTCTCTCAAGTATCATCAGCAAAAGATGCAAAACTTGTTC
ATCGGCCTCTTCTTTCACAGTTGAATCAGGCGTTCTTGGATTGGAAGAAATGCGGTTCTC
TGAGAAAAATCAACTGGAAGAAAATCAAGATGAGAATTGCTCGGGTGTTCCTACTGT
GAACAAGTATCCCTTACCAACGAAACAGGTAAGTGGCGACATGAAACAAGTAAGACCTC
AACTGTGGACAACGCGGTTCAAGATGTTCCCAAGAAGACAAAGACAAAGATGGTAACGA
TGGTACTACTGTGCACAGCATGCAAACTACCTTGGCATTTCACGCAGATATTGTGAA
CGGAATATAGCAAAATGCCCTCAAAATCATCCCTCAGGTATGGTATCTCAAGACTTCAT
GTTTCATCCTATGAGAGAAGAACTCACGGGCACGCAAACTTCAAGCTACAACAGCATC
TGCTACTACTACAGCTTCTCATCAAGCGTTTCCAGCTTGTCAATTACAGGATGATTACCG
TTCGTTTCTCCAGATATCATCTACTTCTCCAATCTTATTATGTCAACTCTCTACAGAA
TCTGTCAGCTCATGCTGCAGCTACATTGCTGCTTCCGCTGCTGCGCTTATGCGAGTGTGCG
GAATCTGGTGATTCAACCCCCAATGAGCTCTTCTCTCCAGTATAACTGCCATTGCG
CGCTGCTACAGTAGCTGCTGCAACTGCTTGGTGGGCTTCTCATGGACTTCTCTCTGTATG
CGTCCAGCTCCAATAACATGTGTTCCATTCTCAACTGTTGCAGTTCCAACCTCCAGCAAT
GACTGAAATGGATACCGTTGAAATACTCAACCGTTTGAGAAACAAACACAGCTCTGCA
AGATCAAACCTTGGCTTCGAAATCTCCAGCTTCATCATCTGATGATTAGATGAGACTGG
AGTAACCAAGCTAAATGCCGACTCAAAACCAATGATGATAAAATTGAGGAGGTTGTTGT
TACTGCCGCTGTGCATGACTCAAACTGCCCAGAGAAAAATCTTGTGGACCGCTCATC
GTGTGGCTCAAATACACCTTCAGGGAGTGACGCAGAACTGATGCATTAGATAAAATGGA
GAAAGATAAAGAGGATGTGAAGGAGACAGATGAGAATCAGCCAGATGTTATTGAGTTAAA
TAACCGTAAGATTAAATGAGAGACAACAACAGCAACAACATGCAACTACTGATTTCGTG
GAAGGAAGTCTCCGAAGAGGGTCGTATAGCGTTTTCAGGCTCTCTTTGCAAGAGAAAGATT
GCCTCAAAGCTTTTCGCTCTCAAGTGGCAGAGAAATGTGAATAGAAAACAAAGTGACAC
GTCAATGCCATTGGCTCCTAATTTCAAAGCCAGGATTCTTGTGCTGCAGACCAAGAAGG
AGTAGTAATGATCGGTGTTGGAACATGCAAGAGTCTTAAACGAGACAGACAGGATTTAA
GCCATACAAGAGATGTTCAATGGAAGTGAAAGAGAGCCAAAGTTGGGAACATAAACAATCA
AAGTGATGAAAAAGTCTGCAAAAGGCTTCGATTGGAAGGAGAAGCTTCTACATGACAGAC
TTGGAGGTAAAAAATACATCCACATTTTATCAATATCTTTAAATCTAGTGTAGTAG
TTTGCTTCTCCAATCTTTATGAAAGAGACTTTTAAATTTCTTCCGAACATTTCTTTGGT
CATGTCAGGTTCTGTACCATATTACCCCATGTCTTGTCTCTGTCTCTGTTGTGTATGC
TACTTGTGGTCTATATGTCATCTGCTACTACTGTTAATTAACCATTAAAGCAATGGATTG
TCTTTA

>G680 Amino Acid Sequence (domain in AA coordinates: 24-70)
MDTNTSGEELLAKARKPYTITKQRRWTEDEHERFLEALRLYGRAWQRIEEHIGTKTAVQ
IRSHAQKFFTKLEKEAEVKGIPVCQALDIEIPPRPKRKPNTPYPRKPGNNGTSSSQVSS
AKDAKLVSSASSSQLNQAFDLLEKMPFSEKSTSGKENQDENCSTVNVKYLPTKQVSG
DIETSKTSTVDNAVQDVPKKNKDKDNDGTTVHSMQNYPHWFHADIVNGNIAKCPQNHP
GMVSDQFMFHPMREETHGHANLQATTASATTTASHQAPPACHSQDDYRSFLOISSTFSNL
IMSTLLQNPAHAHAATFAASVWPYASVGNSSGDSSTPMSSPPSITAIAAATVAAATAWWA
SHGLLPVCAPAPITCVFSTVAVPTPAMTEMDTVENTQPFQKQNTALQDQTLASKSPASS
SDDSDETGVTKLNADSKTNDKIEEVVTAHVDSNTAQKKNLVDRSSCGSNTPSGSDAE
TDALDKMEKDKEDVKETDENQPDVIELNNRKIKMRDNNNNNATTDSWKEVSEEGRIAFQ
ALFARERLPQSFSPPQVAENVNRKQSDTSMPLAPNFKSQSDCAADQEGVVMIGVTCKSL
KTRQTGFKPKYKRCSEMEVKESQVGNINNQSDEKVKRLRLEGEAST*

>G867 (64..1098)

CACAACACAAACACATTTCTGTTTTCTCCATTGTTTCAAACCATAAAAAAACACAGAT
TAAATGGAATCGAGTAGCGTTGATGAGAGTACTACAAGTACAGGTTCCATCTGTGAAACC
CCGGCGATAACTCCGGCGAAAAAGTCGTCGGTAGGTAACCTATACAGGATGGGAAGCGGA
TCAAGCGTTGTGTTAGATTAGAGAACGGCGTAGAAGCTGAATCTAGGAAGCTTCCGTCG
TCAAAATACAAAGGTGTGGTGCCACAACCAACGGAAGATGGGGAGCTCAGATTTACGAG
AAACACCAGCGCGTGTGGCTCGGGACATTCAACGAAGAAGACGAAGCCGCTCGTGCCTAC
GACGTCGCGGTTACAGGTTCCGTCGCGGTGACGCGCTCAGAAATTTCAAAGACGTGAAG
ATGGACGAAGACGAGGTCGATTTCTGAATTCTCATTGAAATCTGAGATCGTTGATATG
TTGAGGAAACATACTTATAACGAAGAGTTAGAGCAGAGTAAACGGCGTCGTAATGGTAAC
GGAAACATGACTAGGACGTTGTTAACGTCGGGGTTGAGTAATGATGGTGTCTTACGACG

GGGTTTAGATCGGCGGAGGCACTGTTTGAGAAAGCGGTAACGCCAAGCGACGTTGGGAAG
CTAAACCGTTTGGTTATACCGAAACATCACGCAGAGAAACATTTCCGTTACCGTCAAGT
AACGTTTCCGTGAAAGGAGTGTTGTTGAACTTTGAGGACGTTAACGGGAAAGTGTGGAGG
TTCCGTTACTCGTATTGGAACAGTAGTCAGAGTTATGTTTGGACTAAAGGTTGGAGCAGG
TTCGTTAAGGAGAAGAATCTACGTGCTGGTGACGTGGTTAGTTTCAGTAGATCTAACGGT
CAGGATCAACAGTTGTACATTGGGTGGAAGTCGAGATCCGGGTCAGATTTAGATGCGGGT
CGGGTTTTGAGATTGTTCCGAGTTAACATTTACCGGAGAGTTCAAGAAACGACGTCGTA
GGAAACAAAAGAGTGAACGATACTGAGATGTTATCGTTGGTGTGTAGCAAGAAGCAACGC
ATCTTTCACGCCTCGTAACAACCTCTTCTCTTTTTTTTTCTTTTGTGTTTAAATAATTT
TTAAAAACTCCATTTTCGTTTCTTTATTTGCATCGGTTTCTTTCTTCTGTTTACCAA
GGTTCATGAGTTGTTTTGTTGTATTGATGAAGTGTAAATTTTATTATAGGATAAATTT
TAAAAA

>G867 Amino Acid Sequence (domain in AA coordinates: 59-124)

MESSSVDESTTSTGSICETPAITPAKKSSVGNLYRMGSGSSVVLDSENGVEAESRKLPS
KYKGVVPQPNGRWGAQIYEKHQRVWLGTFFNEEDEAARAYDVAVHRFRRRDAVTNFKDVKM
DEDEVDFLNSHSEIIVDMLRKHTYNEELEQSKRRRNGNGNMTRTLTSLGSLNDGVSTTG
FRSAEALFEKAVTPSDVGLNRLVLPKHAEKHFPLPSSNVSVKGVLLNFEDVNGKVWRF
RYSYWNSSQSYVLTGKWSRFVEKNLRAGDVVSFRRSNGQDQQLYIGWKSRSRSGDLDA
VLRFLGVNISPESRNDVVGKRVNDTEMLSLVCSKKQRIHAS*

>G956 (1..840)

ATGGAGGAGACAGAAAAGAATAAGGCGCAGCATAAGTATGGTTGAGGCTAATCTACCTCCT
GGTTTTAGATTCCATCCTAGAGACGACGAGCTCGTCTGTGACTACTTAATGAGAAGAACC
GTTTCGCAGCCTCTATCAACAGTTGTCTTGATCGACGTCGATCTTAACAAATGCGAGCCT
TGGGACATTCTCAAACGGCGAGAGTGGGAGGGAAGAATGGTACTTTTACAGCCAAAAA
GACCGTAAATACGCAACAGGCTACAGAACAAACCGGGCTACGGCCACCGGTTATTGGAAA
GCCACCGGGAAGATAGAGCAATCCAAAGAAACGGTGGTCTTGTGGGTATGAGAAAGACA
CTTGTGTTTTACCGAGGTCGATCCCTAAAGGTCGTAAACTGATTGGGTCATGCATGAG
TTTCGTCTCCAAGAAAACCTTCTTCACTCCCTAATTCTCTCGAGGAAGAGTGGGTA
TTGTGTAGAGTTTTCCACAAGAACAGCAACCGAGCTGATATAGACGACATCAAGGAGC
TGCTCTGATGCAACAGCTTCTGCATTCATGGACTCTTACATCAACTTCGACCATCATCAC
ATCATCAATCAGCATGTACCCTGCTTCTCCAATAATTTGTCACATAACCAACCAACCAA
TCCGGTTTAAATCTCCAAGAACTCCAGCCCATTTGTTTAAATGCTTCCCTGATCAAATGATT
CTCAGAACTTTGCTAAGTCAACTCACAAAAAAGTCGAAGAATCACAGAGTCGTGGAGAC
GGAAGCTCAGAGAGCCAATTGACCGACATTGGCATCCCAAGCCATGCATGGAATTACTGA
>G956 Amino Acid Sequence (domain in AA coordinates: TBD)

MEETKKNKSISMVEANLPPGFRFHRPRDELVC DYLMRRRTVRSLYQPVVLDVLDLNKCEP
WDIPQTARVGGKEWYFYSQKDRKYATGYRTNRTATGYWKATGKDRAIQRNGLVGMKRT
LVFYRGRSPKGRKTDWMHEFRLQGKLLHHSNPSLEEELVLCRVFHKNSNGADIDITRS
CSDATASAFMDSYINFDDHHIINQHVPCFSNNLSHNQTNQSLISKNSSPLFNASPDQMI
LRTLQLTKKVEESQSRGDSSESQSLTDIGIPSHAWNY*

>G996 (53..1063)

CGATCGATCTTGAATTGATTTCTTTGTAGTATTTTATTACATATATATATAGATGGGAAG
ACATTCATGTTGTTACAAACAGAACTGAGGAAAGGACTTTGGTCTCCTGAAGAAGATGA
GAAGCTTCTTTCGTTACATCACTAAGTATGGTCATGGTTGCTGGAGCTCTGTCCCTAAACA
AGCTGGTTTACAGAGATGTGAAAAAGTTGTAGATTAGATGGATAAATTATTAAAGACC
AGATTTGAAGAGAGGAGCATTTTCTCAAGATGAAGAAATCTCATTATTGAACCTTCATGC
CGTTCTTGGCAATAGATGGTCTCAGATAGCTGCACAGCTTCTGGAAGAACCACAAATGA
AATCAAGAATCTTTGGAATCTTGTGTTGAAGAAGAAATTGAGGCTGAGAGGAATTGACCC
GGTTACACACAAGCTCTTAACCGAAATCGAAACCGGTACAGATGACAAAACAAAACCGGT
TGAGAAGAGTCAACAGACCTACCTCGTTGAGACTGATGGCTCCTCTAGTACCACTACTTG
TAGTACTAACCACAAACAACTGATCATCTTTATACCGGAAATTTCCGGTTTTCAACG
GTTAAGTCTAGAAAACGGTTCAAGAATCGCAGCCGGTTCTGACCTCGGTATCTGGATTCC
CCAAACCGGAAGAAACCATCATCATCATGTGATGAAACCATCCCTAGTGCAGTGGTACT
ACCCGGTTCAATGTTCTCATCCGGTTTAACCGGTTATAGATCCTCCAATCTCGGTTTAAAT
TGAATTGGAAAACCTATTCTCAACCGGGCCAATGATGACAGAGCATCAGCAAATTCAGA
GAGTAACCTACAACAATTCAACATTTCTTGGAAATGGGAATCTGAATTGGGGATTACAAT
GGAGGAAAATCAAAATCCATTACAAATATCGAATCATTCAAATTCGTCTTATACAGTGA

TATAAAATCAGAGACCAATTTTTTTGGCACAGAGGCTACAAATGTTGGTATGTGGCCATG
TAACCAGCTTCAGCCTCAGCAACATGCATATGGCCATATATAAATCTTCTGTATATTAT
AA

>G996 Amino Acid Sequence (domain in AA coordinates: 14-114)
MGRHSCCYKQKLRKGLWSPEEDEKLLRYITKYGHGCWSSVPKQAGLQRCGKSCRLRWINY
LRPDLKRGAFSQDEENLIIELHAVLGNRWSQIAAQLPGRTDNEIKNLWNSCLKKKLRRLRG
IDPVTHKLLTEIETGTDKTKPVEKSQQTLYLVETDSSSTTTCSTNQNNNDHLYTGNFG
FQRLSLENGSRIAAGSDLGWIWPQTGRNHHHHVDETIPSAVVLPGSMFSSGLTGYRSSNL
GLIELENSFSTGPMTEHQIQESNYNNSTFFNGNLNLWGLTMEENQNPFITSNHSNSSL
YSDIKSETNFFGTEATNVGMWPCNQLQPQGHAYGHI*

>G1946 (90..1547)
TCTCACCTATTGTAAAAATCACCAGTTTCGTATATAAAACCCCTAATTTTCTCAAAATTC
CAAATATTGACTTGAATCAAAAATCCGAATGGATGTGAGCAAAGTAACCACAAGCGACG
GCGGAGGAGATTCAATGGAGACTAAGCCATCTCCTCAACCTCAGCCTGCGGCGATTCTAA
GTTCAAACGCGCCTCCTCCGTTTCTGAGCAAGACCTATGATATGGTTGATGATCACAAATA
CAGATTGATTTGCTCTTGGAGTGCTAATAACAACAGTTTATCGTTTGGAAACCAACCGG
AGTTCGCTCGCGATCTTCTTCTAAGAACTTTAAGCATAATAATTTCTCCAGCTTCGTTA
GACAGCTTAATACCTATGGTTTTCAGGAAGGTTGACCCAGATAGATGGGAATTTGCGAATG
AAGGTTTTTTAAGAGGTCAGAAGCACTTGCTACAATCAATAACTAGGCGAAAACCTGCCC
ATGGACAGGGACAGGGACATCAGCGATCTCAGCACTCGAATGGACAGAACTCATCTGTTA
GCGCATGTGTTGAAGTTGGCAAATTTGGTCTCGAAGAAGAAGTTGAAAGGCTTAAAAGAG
ATAAGAACGTCCTTATGCAAGAACTCGTCAGATTAAGACAGCAGCAACAGTCCACTGATA
ACCAACTTCAAACGATGGTTTCAGCGTCTCCAGGGCATGGAGAATCGGCAACAACAATTAA
TGTCATTCTTGCAAAGGCGAGTACAAAGCCCTCATTTTCTATCTCAATTCTTACAGCAGC
AGAATCAGCAAAACGAGAGTAATAGGCGCATCAGTGATACCAGTAAGAAGCGGAGATTCA
AGCGAGACGGCATTGTCCGTAATAATGATTCTGCTACTCCTGATGGACAGATAGTGAAGT
ATCAACCTCCAATGCACGAGCAAGCCAAAGCAATGTTTAAACAGCTTATGAAGATGGAAC
CTTACAAAACCGGCGATGATGGTTTCTTCTAGGTAATGGTACGTCTACTACCGAGGGAA
CAGAGATGGAGACTTCATCAAACCAAGTATCGGGTATAACTCTTAAGGAAATGCCTACAG
CTTCTGAGATACAGTCATCATCAACCAATTGAAACAACCTCTGAAAATGTTTCGGCAGCAT
CAGAAACAACCGAGAAGTGTATTCCTTACCTGATGATCTAACTCTTCCCGACTTCACTC
ATATGCTACCGGAAAATAATTCAGAGAAGCCTCCAGAGAGTTTCATGGAAACCAAACCTGG
GAGGTTCTAGTCCATTACTAGATCCAGATCTGTTGATCGATGATTCTTTGTCCTTCGACA
TTGACGACTTTCCAATGGATTCTGATATAGACCCTGTTGATTACGGTTTACTCGAACGCT
TACTCATGTCAAGCCCGGTTCCAGATAATATGGATTCAACACCAGTGGACAAATGAAACAG
AGCAGGAACAAAATGGATGGGACAAAACCTAAGCATATGGATAATCTGACTCAACAGATGG
GTCTCTCTCTCTGAAACCTTAGATCTCTCAAGGCAAAATCCTTGATTTTGGGAGTTTT
TAAAGCTTTTGGAGTAACACAGTCCCTGAGAGCAGCATATTTCAT

>G1946 Amino Acid Sequence (domain in AA coordinates: 32-130)
MDVSKVTTSDGGGDSMETKPSQPQPAAILSSNAPPPFLSKTYDMVDDHNTDSIVSWSAN
NNSFIVWKPEFARDLLPKNFKHNMFSSFVRQLNTYGFRRKVDPRWEFANEGFLRGQKHL
LQSITRRKPAHQGGQGHQRSQHSNGQNSSVSACVEVGKFGLEEEVERLKRDKNVLMQELV
RLRQQQSTDNQLQTMVQRLQGMENRQQQLMSFLAKAVQSPHFLSQFLQQNQNNESNRR
ISDTSKKRRFRKRDGIVRNDSATPDGQIVKYQPPMHEQAKAMFKQLMKMEPYKTGDDGFL
LGNGTSTTEGTEMETSSNQVSGITLKEMPTASEIQSSSPIETTPENVSAASEATENCIPS
PDDLTLPDFTHMLPENNSEKPPESFMEPNLGGSSPLLDPLLIDDSLSFDIDDFPMDSDI
DPVDYGLLERLLMSSPVPDNDSTPVDNETEQEQNGWDKTKHMDNLTQQMGLLSPETLDL
SRQNP*

>G217 (84..2618)
cttcgttcttaccgaggtccacgagcattagcttcagagacctgaattggagtgcggtt
ggatcaaaaacagttgagcgaagatgaggattatgattaagggaggtgtttgaagaaca
ccgaagatgagattctcaagccgctgatgaagtatggtaagaaccaatgggctcga
tctcgtctcttcgttcgtaagtctgctaaacagtgtaagctcgctggtacgagtggt
tcgatccatctatcaaaaagactgaatggaccagagaagaagatgagaagcttctacatc
ttgctaaactctgcctactcaatggagaactattgctcctattgtgggtcgtaaccat
ctcaatgtccttgagaggtatgagaagctccttgatgcagcatgcactaaggatgaaaatt
atgatgcagcggtatgacacgaaaattacgtcctggtgagattgatccgaaccagaag

caaagcctgctcgtcctgatccggtagacatggacgaagatgagaaagaaatgctttctg
aagcaagagctagatttggttaacacgaggggaaagaaggctaaaagaaaagctagagaaa
aacaacttgaggaagctagaaggcttgcttctctgcaaaaaagaagagaactaaaagcag
ctgggattgatggaaggcatagggaaagaaagagaaagggaatcgactataatgcagaaa
ttccttttgaaaagaggcaccctgcggttcttatgatactgaggatgaagatcgctctg
ctgatcaagtaaaatttccaactaccattgaagaacttgaaggaaaaagaagagctgatg
tagaagcacatttacgcaacaagatgttgcaaggaaataaaattgctcagagacaggatg
ctccagcagctatattgcaagcaacaagctgaatgatccggaagttgttaggaagaggt
caaagctgatgttaccaccaccgcagatttccagaccacgagctagaagaaattgctaaga
tggtctatgccagtaccttcttgccgagaatgaggagctaacagaaggcagtgctgcta
ctcgtgcacttttgcaaaattactcacaacaccaaggcaaggaaatgacacccatgagga
cacctcaaagaactcctgctggttaaaggtgatgctattatgatggaagcagaaaacctgg
ccagattaagagactctcagacacctttgctaggaggagaaaaatcctgagttgcaccctt
ctgacttcaactggggtcactccgagaaagaaggagattcaaacgcctaatacaatgttga
cccttcaatgactcctggtggtgctggtcttactccaagaattggcttgacgccatcaa
ggatgggtccttcttttctatgacacccaaaggactcccttcagggatgaacttcaca
ttaacgaagacatggacatgcaaaagtgcacaaacttgagaggcagagacgagaggaag
ctagaaggagtttacgctctggtttgactgggttctcctcagccaaagaacgagtaccaa
tagttgcacacctcctcctgaggaaagtgaagagccagaagagaaaaattgaggaagaca
tgtcagacaggatagcgagggaaaggcggaggaagaagcaagacaacaggcattgctta
agaagagatccaaggtccttgagagagatcttctcagacccccagctgcttcattggcag
taattaggaactcgttgctttcagctgatggagacaaaagttctgttctcctcactc
cgattgaggttgagataaaaatggtaagagaggagcttctacagttgctggagcatgata
atgcaaagtatccgcttgatgacaaagctgagaagaagaaggagccaagaaccgtacca
accgttctgcttctcaggttcttgcaattgacgattttgatgaaaatgagctccaagagg
ctgacaaaaatgataaaggaggagggaagttctgtgtgtgtcaatgggacatgagaaca
agacacttgatgattttgtagaagctcacaacacatgcgtgaatgatctcatgtatttcc
ccactcgaagcgcttacgagctctcaagtgttgctgggaacgcggacaaagttgcagctt
ttcaggaggagatggagaatgtgagaaaaaagatggaggaggatgagaagaaggcagaac
acatgaaggccaagtacaaaacttatacaaagggtcatgagaggagggcagagaccgtgt
ggacccaaatagaggcgacattgaagcaggctgagattggtggaaagagaagtagagtgt
ttaaagcattgaaggaggcaagaagagatggctgcatcttttaggaaaaagaatttgcaag
aggaagtataaagcaaaaggaaacagagagtaaaactgcagactcgctatgggaatatgt
tggcaatgggtgaaaaagcagaggagataaatgggtcggtttccgagcacaggcattgaaga
aacaagaggatgttgaagatttccacaaactgaaagaagctaaagctagccactggagagg
aagaggacatagccatagccatggaagcttctgcataaaaacttgagttttgtattgctt
acaagttttaaggagacgtagcttgactttgtattggttaagtttttttaatatgagtc
gactttgtaaaaaggttatgatataattctctgtttgtatgctttgcaagagtcaagaaat
ttgaatgcttcaggatcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa
>G217 Amino Acid Sequence (conserved domain in AA coordinates: 8-67)
MRIMIKGGVWKNTEDEILKAAVMKYGKNQWARISSLLVRKSAKQCKARWYEWLDPSIKKT
EWTREDEKILLHLAKLLPTQWRTIPIVGRTPSQCLERYEKLLDAACTKDENVDAADDPR
KLRPGEIDPNPEAKPARPDPVDMDEDEKEMLSEARARLANTRGKKAKRKAREKQLEEARR
LASLQKRRELKAAGIDGRHRKRKRKGIDYNAEIPFEKRAPAGFYDTADEDPRADQVKFPT
TIELEGKRRADEAHLRKQDVARNKIAQRQDAPAILQANKLNDPEVVRKRSKLMPLPP
QISDHELEELIAKMGYASDLLAENEELTEGSAATRALLANYSQTPRQGMTPMRTPORTPAG
KGDAIMMEAEENLARIIRDSQTPLLGGENPELHPSDFTGVTPRKKEIQTPNPMLTPSMTPGG
AGLTPRIGLTPSRDSSFSMTPKGTPFRDELHINEDMDMQQSAKLERQRREEARRSLRSG
LTGLPQPKNEYQIVAPPPPEESEPEEKIEEDMSDRIAREKAEERQALLKKRSKVLQ
RDLPRPPAASLAVIRNSLLSADGDKSSVVPPTPIEVADKMVREELLQLLEHDNAKYPLDD
KAEEKKGAKNRTNRSASQVLAIDDFDENELQEADKMIKEEGKFLCVSMGHENKTLDDFVE
AHNTCVNDLMPFPTRSAYELSSVAGNADKVAAPQEEMENVRKMEEDEKKAHEMKAKYKT
YTKGHERRAETVWTQIEATLKQAEIGTEVECFKALKRQEEMAASFRKKNLQEEVIKQKE
TESKLQTRYGNMLAMVEKAEIIMVGFAQALKKQEDVEDSHKLKEAKLATGEEEDIAIAM
EASA*
>G2192 (92..2971)
CGGAAAGAGATCAACCAACGATAGAGGAGAAGAAGAACTGCATACGCAAAAAAAGTTTC

CCGGGAAAATTCAGAAACTGCTTTGGAAAAATGTGCGAGCCCGATGATAATTCCGCTAG
AAACGGCGTCACTACTCAACCTTCGAGGTCAAGGGAGCTTCTAATGGATGTTGACGACTT
AGATCTTGACGGTTCATGGCCACTAGATCAAATCCCTTACTTATCCTCATCGAATCGCAT
GATTTCTCCGATTTTTGTCTCCTCTTCTCTGAGCAGCCTTGCTCGCCTCTCTGGGCTTT
CTCCGACGGTGGAGGAAATGGTTTTACCACGCAACCTCCGGTGGCGATGATGAGAAGAT
CAGCTCTGTCTCCGGTGTTCTTCTTCCGTCTCGCCGAGTATCCTCTCTTCTCCCTTA
CTCTTCTCCATCAGCAGCTGAGAACACAACAGAGAAGCATAACAGTTTCCAGTTTCCGTC
TCCATTGATGAGCCTAGTCCCACCAGAGAACACAGACAATACTGTGTGATCAAAGAGAG
GATGACTCAGGCGCTTCGATACTTCAAAGAATCAACCGAACACACGTTTTGGCTCAGGT
CTGGGCTCCTGTGAGAAAGAATGGTCGTGATTTGCTGACGACTTTGGGTCAACCTTTTGT
TCTTAATCCTAATGGTAATGGGCTTAATCAATACAGGATGATCTCTCTCACATATATGTT
TTCTGTGGATAGTGAAAGTGACGTAGAGCTCGGACTCCCGGGTCGAGTTTTCCGTCAGAA
ATTGCCTGAATGGACTCCAAATGTTTCAGTACTATTCCAGCAAAGAATTCTCGCGGCTTGA
TCACGCCCTTGCACTACAACGTGCGTGGTACACTGGCCTTGCTGTCTTTAATCCCTCTGG
TCAGTCCCTGCATAGGTGTTGTGGAACCTTATAATGACCTCAGAGAAGATTCACTATGCACC
CGAAGTGGACAAAGTTTGCAAAGCCCTTGAGGCGGTAAATCTGAAAAGCTCGGAAATACT
TGATCACCACAAACACAGATATGCAATGAGAGTCGCCAAAACGCGCTTGCTGAGATTCT
CGAAGTGCTGACAGTTGTATGTGAGACCCATAACTTGCCCTCTCGCTCAGACTTGGGTTC
ATGTCAGCATGGGAGCGTTCTTGCCAATGGTGGCGGTCTAAAGAAAACTGCACCAGCTT
TGACGGTAGCTGCATGGGTCAAATCTGCATGTCTACAACCGACATGGCCTGCTATGTCGT
GGATGCTCATGTCTGGGGCTTTAGAGATGCCTGTCTTGAAACACCATCTCCAGAAAGGCCA
GGGAGTCGCTGGACGAGCTTTTCTCAATGGTGGCTCATGTTTCTGCAGAGACATCACCAA
GTTCTGCAAAACGCAGTACCCACTAGTCCATTATGCGCTCATGTTCAAGTTGACCACTTG
TTTTGCAATATCTCTCCAGAGCTTTACACGGGCGACGACAGTTACATTCTTGAATTTTT
TCTTCTCCTTCGAGTATAACAGACGACCAAGAGCAAGATTGCTGTTGGGTTCTATTTGGT
GACAATGAAAGAACATTTTCAGAGTCTGAGGGTTGCATCTGGGGTTGACTTTGGTGAAGA
TGACGACAAATTTGCTTTTCGAGATCATCCAAGCATTACCGGACAAGAAGGTTTCATTCAA
AATAGAATCCATTTCGAGTTCCCTTTTCTGGTTTTAAGTCAAATGCAACAGAGACGATGTT
GATTCCTCAGCCTGTGGTTTCAGTCTTCTGATCCAGTAAATGAGAAAATCAACGTGGCCAC
TGTTAACGGTGTGGTTAAGGAGAAGAAGAAAACAGAGAAAAGCGTGGGAAGACTGAGAA
AACAATCAGTCTAGATGTACTTCAGCAGTATTTCACTGGGAAGTCTCAAAGACGCTGCAAA
GAGCCTAGGAGTTTGCCCGACGACAATGAAGCGAATTTGCAGGCAACACGGAATCTCGCG
GTGGCCATCGAGTATAACAGAAAGTGAATCGTTCAATCACAAGCTGAAACGAGTCAT
CGAATCTGTTCAAGGTAAGTACTGATGGAGCCCTCGACCTGACTTCCATGGCCGTTAGTTCCAT
CCCTTGGACACACGGTCAAACATCAGCACAGCCACTAACTCACCCAATGGTTCCAAACC
ACCTGAGCTACCAAACACCAATAATTCACCTAACCATTTGGTCAAGTGATCACAGTCCGAA
CGAGCCAAATGGTTTCGCTGAGTTACCACCAAGCAATGGTCAAGCGATCACGAACGGT
GGATGAGAGCGCTGGGACTCCAACTCTCATGGCTCATGTGACGGTAACCAATTAGATGA
ACCGAAAGTCCCAAATCAAGATCCGCTCTTACGGTTGGTGGATCACCGGGCTCCTTTT
TCCACCTTATCTAGAGATCATGATGTATCTGCAGCTTCTTCGCAATGCCGAACAGGCT
TCTTGGTTCTATAGACCATTTCCGAGGAATGCTCATTGAAGACGCTGGAAGTTCAAAGA
TCTGAGAAATCTCTGCCCCACTGCAGCATTTGACGATAAGTTTCAAGACACAACTGGAT
GAACAATGATAATAATAGCAACAACAATTATACGCTCCCCCAAAGGAAGAGGCCATTGC
AAATGTTGCATGCGAACCATCAGGCTCAGAAATGAGAACGGTAACAATCAAAGCAAGTTA
CAAAGACGACATAATACGGTTTCAAGATATCCTCGGGTTTCAAGTATAATGGAATTGAAGGA
TGAAGTGGCTAAGAGGCTGAAAGTTGATGCAGGAACGTTTCGATATCAAGTATCTTGACGA
TGATAACGAATGGGTTTTAATAGCTTGTGATGCTGATCTTCAAGAATGTCTCGAGATCCC
TAGATCCTCCCGCAEGAAAATCGTAAGGCTCTTAGTTTCATGATGTAACGACAAATCTAGG
GAGCTCCTGCGAGAGCACTGGAGAATTGTGACCTGATAATTCATTGCAACTCTTTGTAA
ATAG

>G2192 Amino Acid Sequence (conserved domain in AA coordinates:600-700)

MCEPDDNSARNGVTTQPSRSRELLMDVDDLDLDGWSPLDQIPYLSSSNRMISPIFVSSSS
EQPCSPWAFSDGGNGFHHATSGGDDEKISSVSGVPSFRLAEYPLFLPYSSPSAAENTT
EKHNSFQFPSPMLSLVPPENTDNYCVIKERMTQALRYFKESTEQHVLAQVWAPVRKNGRD
LLTTLGQPFVLNPNNGNLNQYRMSLTVMFSVDSESDVELGLPGRVFRQKLPEWTFPNVQY
YSSKEFSRLDHALHYNVRGTLALPVFNPSGQSCIGVVELIMTSEKIHYAPEVDKVCALAE
AVNLKSSEILDHQTQICNESRQNALAEILEVLTVCETHNLPLAQTWVPCQHGSVLANG

GGGLKKNCTSFDGSCMGQICMSTTDMACYVVDHAVHWGFRDACLEHHLQKGQGVAGRAFLNG
GSCFCRDITKFKCTQYPLVHYALMFKLTTCAISLQSSYTGDDSYILEFFLPSSITDDQE
QDLLLGSILVTMKEHFQSLRVASGVDFGEDDDKLSFEIIQALPKKVHVKIESIRVPFSG
FKSNATETMLIPQPVVQSSDPVNEKINVATVNGVVKEKKTEKKRGKTEKTISLDVLQOY
FTGSLKDAAKSLGVCPTTMKRICRQHGISRWPSRKIKKVNRSITKLKRVIESVQGTGGGL
DLTSMVSSIPWTHGQTSAQPLNSPNGSKPPELPNTNNSPNHWSSDHSPNEPNGSPPELPP
SNGHKRSRTVDESAGTPTSHGSCDGNQLDEPKVPNQDPLFTVGGSPGLLFPYSRDHDVS
AASFAMPNRLGSLDHFRGMLIEDAGSSKDLRNLCPATAAFDDKFQDTNWMNNDNNSNNNL
YAPPKEEAIANVACEPSGSEMRTVTIKASYKDDIIRFRISSGSGIMELKDEVAKRLKVDA
GTFDIKYLDDDDNEWVLIACDADLQECLEIPRSSRTKIVRLLVHDVTNGLSSCESTGEL*

>G504 (69..1040)

CGTCGACCTCTTGACGATCATGAGACTGATTTCTGTGAAAATATCGTCATTATATCAAATT
AGAAGTTGATGAAAAACATGGGGGATTCGAGCATAGGGCCGGGCCATCCGCATCTCCCTC
CCGGGTTTCGGTTTTCACCCGACTGATGAGGAAGTAGTATTACCTCAAGAAGAAAG
CAGATTCTGTTCCACTTCCAGTCTCAATCATCGCAGAGATTGATCTTTACAAGTTTGATC
CTTGGGAGCTTCCAAGCAAGGCGAGTTTTGGAGAGCACGAGTGGTACTTCTTTAGTCCTC
GGGATCGGAAGTATCCAAATGGGGTTAGGCCAAACCGGGCAGCAACTTCCGGTTATTGGA
AAGCAACGGGAACCGATAAACCGATATTTACGTGCAATAGTCACAAGGTTGGTGTCAAGA
AAGCGCTTGTTTTTACGGTGGAAAGCCTCCTAAAGGGATAAAAACAGATTGGATCATGC
ATGAATATCGCCTCACTGATGGTAACCTTAGCACTGCGGCTAAGCCGCTGACTTAACCA
CGACAAGGAAAACTCACTACGGCTAGACGATTGGGTTCTATGTAGGATCTATAAGAAGA
ATAGTTCACAAAGACCAACAATGGAGAGAGTATTACTTAGAGAGGATCTAATGGAAGGCA
TGCTCTCAAAATCATCTGCTAATTCTTCTTCTACATCAGTACTAGACAACAACGACAACA
ATAATAACAATAACGAAGAACACTTTTTCGACGGTATGGTCGTTTCTTCAGACAAACGTT
CCTTGTGTGGTCAATACCGAATGGGCCACGAGGCTCAGGATCATCTTCATTCCGATCTT
TCTTATCGAGCAAGAGTTTCATCATACAGGTGATCTCAACAATGATAACTACAATGTCT
CTTTTGTTCGATGCTTAGTGAGATTCTCAGAGTTCCGGGTTTCATGCAAATGGTGTTA
TGGATACGACGTCGTCTCTAGCTGATCATGGGGTTTTAAGACAGGCGTTTCAGCTTCCTA
ACATGAAGTGGCACTCATAATCTATATAGATATATATGTGTGTATCATATATGTATCTAT
GCAGGCCTAATATAGTTTACACATAAATCATCTGGGGCGGCCGCT

>G504 Amino Acid Sequence (domain in AA coordinates: TBD)

MENMGDSSIGPHPLPPGFRFHPTDEELVVHYLKKKADSVPLPVSIIEIDLYKFDPW
LPSKASFGHEWYFFSPDRKYPNGVRPNRAATSGYWKATGTDKPIFTCNHVKGVKKAL
VFYGGKPPKGIKTDWIMHEYRLTDGNLSTAAKPPDLTTTRKNSLRLLDDWVLCRIYKKNSS
QRPTMERVLLREDLMEGLMSKSSANSSSTSVLDNNDNNNNNNEEHFFDGMVVSDDKRS
GQYRMGHEASGSSSFGSFLSSKRFHHTGDLNNDNYNVSVFVSMLEIPQSSGFHANGVMDT
TSSLADHGVLRQAFQLPNMNWHS*

>G622 (248..2620)

TCTTTCTTTCTCAATTCGCCGTCAAAATCTTCTCTTTCTTTCCCCCGCGGTCCTTCA
CCAATCCTCTGATCTCTACACACGAACCTTTGATTTTGACCAACGTCGATGCATGTTT
ATGACTAGTCTCTTCTCAATCTTCAATTTTCATCAATTCACGTCGATTTCTGATCCGAT
TCGTTGTTCTAGCTCTTTGTGTGGTGTAGGGTTTTAAGATTTTGAATTGGGGTTTGA
GTTTGTGATGTTTGAAGTCAAAATGGGGTCAAGATGTGCATGAACGCTTCATGTGTGAC
GACTTCTACTGTTGAATGGAAGAAAGGTTGGCCTCTTCGATCTGGTCTTCTCGCTGATCT
CTGTTATCGTTGCGGATCTGCGTATGAGAGTTCTCTATTCTGTGAACAATTTTATAAGGA
CCAATCTGGTTGGAGGGAATGCTATTTGTGTAGCAAGAGACTACATTGTGGATGCATTGC
TTCTAAGGTAACGATTGAGTTAATGGACTATGGTGGTGTGGTTGTAGTACATGTGCTTG
CTGCCATCAACTCAATTTGAACACAAGGGGTGAGAATCCAGGTGTTTTTAGCAGATTGCC
AATGAAAACGTTAGCTGATAGGCAACATGTAAATGGCGAAAGCGGAGGAAGAAACGAAGG
CGATCTCTTTTCTCAGCCACTAGTCATGGGCGGAGATAAAGGGAAGAGTTTCATGCCTCA
CCGTGGGTTTGGTAAGCTAATGAGTCCAGAAAGTACAACCACCGGCATAGGCTGGATGC
TGCTGGGGAAATGCATGAATCATCACCTTTACAGCCATCTTTAAATATGGGTTTGGCTGT
GAATCCGTTTAGCCCATCTTTTGAACCGAGGCTGTGAGGGGAATGAAACACATCAGTCC
TTCTCAGTCCAACATGGTCCATTGCTCTGCTTCTAATATACTGCAAAAGCCATCAAGACC
TGCTATTTCAACTCCTCTGTGGCTAGTAAATCCGCTCAGGCGGGATTGGAAGGCCCTCC
TGTCGAAGGGCGAGGAGAGGAGGCACTTGCTTCCGCGGTATTGGCCAAAATATACGGATAA
AGAGGTTTCAGCAGATCTCTGGAAATTTGAATTTGAACATTGTACCTCTCTTTGAGAAAAC

TCTTAGTGCCAGTGATGCTGGTCGCATTGGTCGTCTAGTTCTTCCAAAAGCCTGTGCAGA
 GGCATATTTTCTCCGATTAGTCAATCCGAAGGCATTCCTTTGAAAATCCAAGATGTGAG
 GGGTAGGGAGTGGACGTTCCAGTTCAGATATTGGCCCAATAACAATAGTAGAATGTATGT
 TTTAGAAGGTGTCACTCCATGCATACAGTCCATGATGCTACAGGCTGGTGATACAGTAAC
 TTTCACTCGGGTTGATCCTGGCGGAAAACTAATCATGGGTTCCAGGAAGGCAGCTAATGC
 TGGAGACATGCAGGGTTGTGGGCTCACCAACGGAACATCAACTGAGGACACATCATCGTC
 TGGTGTAAACAGAAAACCCACCCTCCATAAATGGTTCTCGTGTATTTCACTAATACCGAA
 AGAGTTGAATGGTATGCCTGAGAATTTGAACAGTGAGACTAACGGGGGCGAGGATAGGTGA
 TGATCCTACACGAGTTAAAGAGAAGAAGAGAACTCGAACCATTGGTGCAAAAAATAAGAG
 ACTTCTTTTGCATAGTGAAGAATCTATGGAGCTGAGACTCACTTGGGAAGAAGCTCAGGA
 CTTGCTTCGTCCCTCTCCTAGTGTAAGCCTACCATCGTTGTCTATGAGGAGCAAGAAAT
 TGAAGAATATGACGAACCTCCTGTCTTTGGAAGAGGACTATAGTCACTACAAAACCTTC
 AGGTGAACAGGAACGATGGGCAACTTGGCAGCAGTCTCTAAATGGAGAAGGTTACCTGT
 AGATGCTCTTCTTTCTTTAAATGGACATGTATAGACAATGTTTGGGATGTGAGTAGGTG
 TTCATGTTCTGCACCGGAGGAGAGTCTGAAGGAACCTGAGAATGTTCTTAAAGTAGGTAG
 AGAGCACAAGAAGAGAAGAACTGGGGAAAGACAGGCAGCACAAAGTCAGCAAGAACCGTG
 TGGTTTGGACGCGACTGGCGAGTGCAGCAGTCTTAGGAGACACAATAGGCGAGCCAGAGGT
 AGCGACCACGACCAGACATCCAAGGCACAGGGCTGGATGCTCTTGCATCGTGTGCATTCA
 GCCACCAAGTGGGAAAGGTAGGCACAAGCCTACATGTGGCTGCACTGTGTGTAGCACCGT
 GAAGAGAAGGTTCAAGACGCTTATGATGAGGAGGAAGAAGAAGCAGTTGGAGCGCGATGT
 AACAGCAGCAGAAGATAAGAAGAAGAAGGACATGGAACCTGGCTGAGTCTGATAAGAGTAA
 GGAGGAGAAGGAAGTGAACACAGCGAGAATAGACCTGAACAGTGATCCATACAATAAAGA
 AGATGTTGAAGCTGTTGCGGTGGAGAAAGAAGAGAGTCGAAAAAGAGCAATAGGACAGTG
 TTCGGGCGTGGTGGCTCAAGACGCCAGTGATGTTTTAGGAGTTACAGAGTTAGAAGGAGA
 GGGTAAGAATGTTCTGTAAGAGCCGAGAGTTTCAAGCTGATATGGAAA

>G622 Amino Acid Sequence (domain in AA coordinates: TBD)

MFEVKMGSKMCMNASCGTTSTVEWKKGWPLRSGLLADLCYRCGSAYESSLFCEQFHKDQS
 GWRECYLCSKRLHCGCIASKVTIELMDYGGVGCSTCACCHQLNLNTRGENPGVFSRLPMK
 TLADRQHVNGESGGRNEGDLFSQPLVMGGDKREEFMPHRGFGKLMSPSTTTGHRLDAAG
 EMHESPLQPSLNMGLAVNPFSPSFATEAVEGMMKHISPSQSNMVHCSASNILQKPSRPAI
 STPPVASKSAQARIGRPPVEGRGRHLLPRYWPKYTDKEVQQISGNLNLNIVPLFEKTL
 ASDAGRIGRLVLPKACAEAYFPPISQSEGIPLKIQDVRGREWTFQFRYWPNNNSRMVLE
 GVTPTCIQSMMLQAGDVTFTSRVDPGGKLIMGSRKAANAGDMQGCGLTNGTSTEDTSSSGV
 TENPPSINGSSCISLIPKELNGMPENLNSETNNGRIGDDPTRVKEKKRTRTIGAKNKRLL
 LHSEESMELRLTWEEAQDLLRSPSVKPTIVVIEEQEIEEYDEPPVFGKRTIVTTKPSGE
 QERWATCDDCSKWRRLPVDALLSFKWTCIDNVWDVSRSCSAPEESLKELENVLKVGREH
 KKRRTGERQAAQSQEQPCGLDALASAAVLGDTIGEPEVATTTTRHPRHRAGCSCIVCIQPP
 SGKGRHKPTCGCTVCSVKRRFKTLMRRKKKQLERDVTAAEDKKKDMELAESDKSKEE
 KEVNTARIDLNSDPYNKEDVEAFAVEKEESRKRAIGQCSGVVAQDASDVLGVTELEGEGK
 NVREEPRVSS*

>G778 (50..1249)

TCTCAATAACACAAAACCTTTTAAACTAGTAAAATACACAGATTTTAGGATGAGCCAATG
 TGTTCCAAACTGTCAATCGATGATACTCCGGCAGCAGCCACCACCCTCCGCTCCAC
 CACAGCCGCGAGACATCCCCATATTAGACTACGAGGTAGCCGAGCTGACGTGGGAGAACGG
 GCAACTAGGCTTGCACGGCTTAGGTCCACCGCAGTGACGGCTTCGTCGACCAAGTACTC
 CACAGGCGCCGGTGGAACTTGGAGTCGATAGTGGAACCAAGCTACTCGCCTCCCTAAGCC
 TAAGCCACCGATGAGCTCGTCCCGTGGTTCCATCATCGCTCCTCCAGGGCCGCGATGGC
 AATGGACGCGCTTGTCCTTGTCTCAACCTAGTACACGAGCAGCAGAGCAAGCCTGGTGG
 CGTTGGCTCCACCCGGGTGGGGTCATGTAGCGATGGTCGTACCATGGGCGGTGGAAAACG
 AGCAAGAGTGGCACCCGAGTGGAGCGGCGGCGGAGTCAGCGGCTGACCATGGACACTTA
 CGACGTAGGTTTCACTCAACATCAATGGGCTCGCACGATAACACAATCGACGATCATGA
 CTCCGTCTGCGCACCGCCACAGATGGAGGACGAAGAAGAGAAGAAAGCCGGAGGAAA
 ATCATCAGTTTCAACCAAGAGAAGCAGAGCTGCTGCTATTTCATAACCAATCCGAACGTAA
 GAGGAGAGATAAAATCAATCAAAGGATGAAGACTTTGCAAAAACCTGGTTCCCAATTCCAG
 CAAGACGGATAAAGCATCTATGTTGGATGAAGTGATAGAGTATTTGAAGCAACTTCAAGC
 ACAAGTGAGCATGATGAGCAGAATGAATATGCCTTCTATGATGCTTCTATGGCCATGCA
 GCAACAACAACAACTACAAATGTCTCTCATGTCCAATCCCATGGGTTTAGGGATGGGCAT

GGGGATGCCCCGCTCTCGGTCTCCTCGACCTTAATTCATGAACCGAGCTGCTGCAAGCGC
 TCCTAATATCCATGCCAACATGATGCCAAACCCATTTTGGCCATGAATTGTCCATCGTG
 GGATGCTTCTTCCAATGACTCTCGATTTTCAGTCTCCTCTCATCCCCGATCCTATGTCTGC
 CTTTCTTGCTGCTCTACTCTAGCCAACGACGATGGAAGCGTATAGCAGGATGGCTACATT
 ATATCAGCAAATGCAACAACAACCTTCTCCTCTCGAATCCAAATGATTATTACTCAA
 ACACCTCTATATAGTTTACGTCTATATATGTGTTAGTCACATACATATATATATATTC
 CATCATAATTATTTATTTATATGTATAGGCTTCTCATGAATTATGATATTATACGTATTA
 CGTAAAAAA

>G778 Amino Acid Sequence (domain in AA coordinates: 220-267)
 MSQCVPNCHIDDTCAAATTTVRSTTAADIPILDYEVAEELTWENGQLGLHGLGPPRVTASS
 TKYSTGAGGTLESIVDQATRLPNPKPTDELVPWFHRRSSRAAMAMDALVPCSNLVHEQQS
 KPGGVGSTRVGS CSDGR TMGGGKRARVAPEWSSGGGSQLTMDTYDVGFSTSTSMGSHDNTI
 DDHDSVCHSRPQMEDEEEKKAGGKSSVSTKRSRAAIIHNQSERKRRDKINQRMKTLQKLV
 PNSSKTDKASMLDEVIEYLKQLQAQVSMMSRMNMPMSMLPAMQQQQQLQMSLMSNPMGL
 GMGMGMPLGLLLDLNSMNRAAASAPNIHANMMPNPFPLMNCPSWDASSNDSRFQSPILPD
 PMSAFLACSTQPTTMEAYSRLMATLYQQMQQLPPPSNPK*

>G791 (173..877)
 TTTTCTTTGGGTGTTCTTCCACCAACGGCAGAAATCGATTCGGCTTAAATCTCCCCCTC
 CTTTCGATCTCTCTGATCGCCGCCGGGAACATTCAATTTCCCGGGAGTTCAACAAAAAA
 AAATCTCCCGTTTTATTTTTCCCTTTTTCACCGGTGGAAGTTTCCGGAGATGGTGTC
 ACCCGAAAACGCTAATTGGATTTGTGACTTGATCGATGCTGATTACGGAAGTTTCAAT
 CCAAGTCTCTGTTCTCTTGGCCTGTTTCAGCAACCTATTGGTGTTTCTTCTAACTCCAG
 TGCTGGAGTTGATGGCTCGGCTGGAACTCAGAAGCTAGCAAAGAACCTGGATCCAAAA
 GAGGGGGAGATGTGAATCATCTCTGCCACTAGCTCGAAAGCATGTAGAGAGAAGCAGCG
 ACGGGACAGGTTGAATGACAAGTTTATGGAATTGGGTGCAATTTTGAGCCTGGAAATCC
 TCCCAAAACAGACAAGGCTGCTATCTTGGTTGATGCTGTCCGCATGGTGACACAGCTACG
 GGGCGAGGCCCAGAAGCTGAAGGACTCCAATTCAAGTCTTCAGGACAAAATCAAAGAGTT
 AAAGACTGAGAAAAACGAGCTGCGAGATGAGAAACAGAGGCTGAAGACAGAGAAAGAAAA
 GCTGGAGCAGCAGCTGAAAGCCATGAATGCTCCTCAACCAAGTTTTTCCAGCCCCACC
 TATGATGCCTACTGCTTTTGGCTTCAGCGCAAGGCCAAGCTCCTGGAAACAAGATGGTGCC
 AATCATCAGTTATCCAGGAGTTGCCATGTGGCAGTTTCATGCCTCCTGCTTCAGTCGATAC
 TTCTCAGGATCATGTCCTTCGTCCTCTGTTGCTTAATCAAGAAAAATCATCAACCGGTT
 TGCTTCTTGCTTCCGCTTAAAGAAAAAGTCTCCATTGTTTGTCTCTCTCTTCTCG
 GCTTTCTTAGTCTTATCCTTTTGTCTTGTGCTGTTATCATCGTAACTGTTATCTGTTGAA
 CAATGATATGACATTGTAACTCCAATTGCTTCGCGCAATGTTATCTATTACATGTAAA
 TTTAAGTAGAGTTTGGCAAAAAAA

>G791 Amino Acid Sequence (domain in AA coordinates: 75-143)
 MVSPENANWICDLIDADYGSFTIQPGFSWVQPIGVSSNSSAGVDGSAGNSEASKEPG
 SKKRGRCESSSATSSKADREKQRRDLNDKFMELGAILEPKNPPKTDKAAAILVDVRMVT
 QLRGEAQKLKDSNSSLQDKIKELKTEKNELRDEKQRLKTEKEKLEQQLKAMNAPQPSFFP
 APPMPTAFASAQGPKNMVPPIISYPGVAMWQFMPPASVDTSQDHVLRPPVA*

>G861 (158..880)
 CTTCTCTCTCTCTCCATCTCTTCTCTTTACTCTCTCTTTAATCATCTCTCATCTTGA
 ATCTTGATCCATCAAAATCAATCCCGTTCTCGAAAGATCCATTAATCAAAACCTAAGC
 TCTCTCTCTTGTCTTAGGGTTTTTTTGTTCGTTGTGATGGCGAGAGAAAAGATTGAGAT
 CAGGAAGATCGACAACGCAACGGCGAGACAAGTGACGTTTTTCGAAACGAAGAAGAGGGCT
 TTTCAAGAAAGCTGAAGAACTCTCCGTTCTCTGCGACGCCGATGTCGCTCTCATCATCTT
 CTCTTCCACCGGAAAAGTTTCGAGTTCTGTAGCTCCAGCATGAAGGAAGTCTTAGAGAG
 GCATAACTTGCGAGTCAAGAACTTGGAGAAGCTTGATCAGCCATCTCTTGAGTTACAGCT
 GGTTGAGAACAGTGATCACGCCCCGAATGAGTAAAGAAATGCGGACAAGAGCCACCGACT
 AAGGCAAATGAGAGGAGAGGAACTTCAAGGACTTGACATTGAAGAGCTTCAGCAGCTAGA
 GAAGGCCCTTGAAGCTGGTTTGACGCGTGTGATTGAAACAAAGAGTGACAAGATTATGAG
 TGAGATCAGCGAAGTTTCAGAAAAAGGGAATGCAATTGATGGATGAGAACAAAGCGTTGAG
 GCAGCAAGGAACGCAACTAACGGAAGAGAACGAGCGACTTGCGCATGCAATATGTAACAA
 TGTGCATGCACACGGTGGTGCTGAATCGGAGAACGCTGCTGTGTACGAGGAAGGACAGTC
 GTCGGAGTCTATTACTAACGCCGGAAGTCTACCGGAGCGCTGTTGACTCCGAGAGCTC
 CGACACTTCCCTTAGGCTCGGCTTACCGTATGGTGTTAGAGATGGAACAATCAAGAA

GTGATGGAGTGAGGAGAGTAATGTAAATCTTTTAACTCGGTAGTAACAAGAGACAATG
TCTAAGTAGTGAATTCTCAAATGTTTGTGTAAGTTTCTGCCTATGGAAGAGGCTTTCATT
TTTATGATTTTCACTATGTATGATCTCTTCACTGCATTTCTGGTTAGTAACGGCTTGT
CACCGATAAATTTCTCGTTATGGAAAGTTAGAATAAAAAAAAAAAAAAAAAAAAAA
>G861 Amino Acid Sequence (domain in AA coordinates: 2-57)
MAREKIQIRKIDNATARQVTFKRRRLFKKAEELSVLCDADVALIIFSSTGKLFECSS
SMKEVLERHNLQSKNLEKLDQPSLELQLVENS DHARMSKEIADKSHRLRQMRGEELQGLD
IEELQQLEKALETGLTRVIETKSDKIMSEISELQKKGMQLMDENKRLRQQGTQLTEENER
LGMQICNNVHAHGAESENAAYVEEGQSSESITNAGNSTGAPVDESSDTSRLGLPYGG
*

>G938 (1..1755)

ATGATGATGTTTAAACGAGATGGGAATGTATGGAAACATGGATTTCTTCTCTTCTCCACA
TCTCTCGATGTGTGCCATTACCACAAGCTGAACAAGAACCTGTAGTTGAAGATGTCCGAC
TACACCGATGATGAGATGGATGTGGATGAGCTTGAGAAGAGGATGTGGAGAGACAAAATG
CGTTTGAACGCTCTCAAGGAGCAACAGAGTAAGTGTAAGAAGGCGTCGATGGTTCGAAA
CAGAGGCAGTCGCAAGAGCAAGCTAGGAGGAAGAAAATGTCTAGAGCCCAAGATGGGATC
TTGAAGTATATGTTGAAGATGATGGAAGTTTGTAAAGCTCAAGGCTTTGTTTATGGTATT
ATTCTTGAGAAGGGTAAGCCTGTGACTGGTGCTTCGGATAATTTGAGGGAATGGTGGAAA
GATAAGGTTAGGTTTGTATCGTAATGGTCCAGCTGCTATTGCTAAGTATCAGTCAGAGAAT
AATATTCTGGAGGGAGTAATGATTGTAACAGCTTGGTTGGTCCAACACCGCATACGCTT
CAGGAGCTTCAGGACACGACTCTTGGTTTCGCTTTTATCGGCTTTGATGCAACATTGTGAT
CCACCGCAGAGACGGTTTCTTTGGAGAAAGGAGTTTCTCCACCTTGGTGGCCTAATGGG
AATGAAGAGTGGTGGCCTCAGCTTGGTTTACCAAATGAGCAAGGTCCTCCTCTTATAAG
AAGCCTCATGATTGAAGAAAGCTTGGAAAGTCGGTGTTTAACTGCGGTGATCAAGCAT
ATGTCGCCGGATATTGCCGAAGATCCGTAAGCTTGTGAGGCAATCAAATGCTTGCAGGAT
AAGATGACGGCGAAAGAGAGTGCTACTTGGCTTGCCATTATTAACCAAGAAGAGGTTGTG
GCTCGGGAGCTTTATCCCGAGTCATGCCCTCCTCTTTCTTCTTCTTCTCATTAGGAAGC
GGGTCGCTTCTCATTAAATGATTGTAGCGAGTATGACGTTGAAGGTTTCGAGAAGGAACAA
CATGGTTTCGATGTGGAAGAGCGGAAACCAGAGATAGTGATGATGCATCCTCTAGCAAGC
TTTGGGGTTGCTAAATGCAACATTTTCCCATAAAGGAGGAGGTCGCCACCACGGTAAAC
TTAGAGTTCACGAGAAAGAGGAAGCAGAAACAATGATATGAATGTTATGGTAATGGACAGA
TCAGCAGGTTACACTTGTGAGAATGGTCAGTGTCTCAGCAAAAATGAATCTTGGATTT
CAAGACAGGAGTTCAAGGGACAACCAACAGATGGTTTGTCCATATAGAGCAATCGTTTA
CGTATGAGCATCCAAGTTTTCATATGGGTGGAATGAAACTAGTAGTTCCTCAGCAACCA
GTCCAACCGATCGACCTATCGGGCGTTGGAGTTCGGGAAAACGGGCAGAAGATGATCACC
GAGCTTATGGCCATGTACGACAGAAATGTCCAAAGCAACCAACGCCTCCTACTTTGATG
GAAAACCAAAGCATGGTCATTGATGCAAAAGCAGCTCAGAATCAGCAGCTGAATTCAAC
AGTGGCAATCAAATGTTTATGCAACAAGGGACGAACAACGGGGTTAACAATCGGTTCCAG
ATGGTGTGTTGATTTCGACACCATTCGATATGGCAGCATTCGATTACAGAGATGATTGGCAA
ACCGGAGCAATGGAAGGAATGGGGAAGCAGCAGCAGCAGCAGCAGCAGCAAGATGTA
TCAATATGGTTCTGA

>G938 Amino Acid Sequence (domain in AA coordinates: 96-104)
MMMFNEMGMYGNMDFSSSTSLDVCPLPQAEQEPVVEDVDYTDDEMDVDELEKRMWRDKM
RLKRLKEQQSKCKEVDGSKQRQSQEQARRKKMSRAQDGILKYMLKMMEVCKAQGFVYGI
IPEKGPVTGASDNLRWWKDKVRFDRNGPAAIAKYQSENNISGGSND CNSLVGPTPHTL
QELQDTTLGSLLSALMQHCDPPQRRFPLEKGVSPWPWPNGNEEWWPQLGLPNEQGPYPYK
KPHDLKKAWKVGVLTAVIKHMSPDIAKIRKLVRQSKCLQDKMTAKESATWLAIINQEEVV
ARELYPESCPLSSSSSLGSGSLLINDCSEYDVEGFEKEQHGF DVEERKPEIVMMHPLAS
FGVAKMQHFPIKEEVATTVNLEFTRKRKQNDNMVMVMDRSAGYTCENGQCPSHKMNLGF
QDRSSRDNHQMVCFYRDNRLAYGASKFHMGMKLVVPQQPVQPIDLSGVGVPENGQKMIT
ELMAMYDRNVQSNQTPPTLMENQSMVIDAKAAQNQQLNFNSGNQMFQQGTNNGVNRRFQ
MVF DSTPFDMAAFDYRDDWQTGAMEGMGKQQQQQQQQQDVS IWF*

>G965 (73..1956)

GATTCTCTGTGTATGTCTGAATCCTTACAGGATCCAAGAGCTTTGGAAAAAGATATAAT
GAATAACAAGATATGGGTTTAGCTACTACAACTTCTTCTATGTCAAGATTATCATCAT
CACCAAGGAATCTTTCTCTCTAATGGATTCCACCGATCATCATCAACCACTCATCAG
GAGGAAGTAGATGAATCCGCCGTCTCTCCGGTGCTCAAATTCGGGTTTATGAAACCGCC

GGAATGTTGTCTGAAATGTTTGCTTACCCTGGCGGAGGTGGCGGCGGTTCCGGTGGAGAG
 ATTCTTGATCAGTCTACTAAACAGTTGCTAGAGCAACAAAACCGTCACAACAACAACAT
 AACTCAACTCTTCATATGTTATTACCAAATCATCATCAAGGTTTTGCTTTCACCGACGAA
 AACACTATGCAGCCGCAGCAACAACAACACTTTACATGGCCATCTTCTCCTCCGATCAT
 CATCAAAACCGAGATATGATCGGAACCGTCCACGTGGAAGGAGGAAGGGTTTGTCTTA
 TCTCTCTCATCTTCATTAGCCGCAGCTAAAGCCGAGGAATATAGAAGCATTATTGTGCA
 GCCGTTGATGGAACCTTCTTCTTCTTAACGCATCCGCTCATCATCAATTCAATCAG
 TTCAAGAATCTTCTTCTTGAGAATTCTTCTTCTCAACATCATCACCATCAAGTTGTGGGA
 CATTTTGGTTCATCATCATCATCTCCCATGGCGGCTTCTTCATCCATTGGAGGGATCTAC
 ACGTTGAGGAATTCGAAATATACGAAACCGGCTCAAGAGTTGTTGGAAGAGTTTTGTAGT
 GTTGAAGAGGACATTTCAAGAAGAACAACCTTAGTAGGAACAACCTCAAACCCTAATACT
 ACCGTTGGAGGAGGAGGCGGAGGGTCTCGTCAATCGGCCGGAACAGCTAATGATAGTCCT
 CCTTTGTCTCCGGCTGATCGGATTGAACATCAAAGAAGAAAAGTCAAGCTACTATCTATG
 CTTGAAGAGGTGGACCGACGGTACAACCACTACTGCGAACAATGCAAATGGTAGTGAAC
 TCATTGACCAAGTAATGGGTACGGCGCGCGGTTCCGTACACGACATTAGCTCAAAAG
 GCAATGTCTAGGCATTTCCGGTGTGTTGAAAGACGCGGTAGCGGTTTCAAGCTTAAACGCAGC
 GTGAGCTTCTAGGGGATAAAGAGGCGGCGAGGGCTGCATCCTCGGGGTTAACCAAGGG
 GAAACGCCGCGATTGCGTTTGTCTAGAGCAGAGTTTGCCTGAGCAACGAGCGTTTCATCAT
 ATGGGTATGATGGAGCAAGAGGCATGGAGACCGCAACGTGGTTTGCCTGAACGCTCCGTT
 AATATCCTTAGAGCTTGGCTATTTCGAGCATTTTCTTAATCCGTACCCAAGCGATGCTGAT
 AAGCACTCTTAGCACGACGACTGGTTTATCCAGAAATCAGGTGTCAAATTGGTTCATA
 AATGCTAGGGTTTCGCCTATGGAACCAATGGTGGAAAGAGATGTATCAACAAGAAGCAAAA
 GAAAGAGAAGAAGCAGAAGAAGAAAATGAAAATCAACAACAACAAGAAGACAGCAACAA
 ACAACAACAACGACACGAAACCCAACAACAATGAAAACAACCTTCACTGTCTATAACCGCA
 CAACTCCAACGACGATGACATCGACACATCACGAAAACGACTCTTCATTCTCTCTTCC
 GTCGCGCGCGCTTCTCACGGCGGTTCAGACGCGTTCACCGTCGCCACGTGTCAGCAAGAC
 GTCAGTGACTTCCACGTCGACGGAGATGGTGTGAACGTCATAAGATTCGGGACCAACAG
 ACTGGTGACGTGTCTCTTACGCTTGGTCTACGCCACTCTGGCAATATTCTGATAAGAAC
 ACTTCTTCTCEGTTAGAGACTTTGGAGATTTTTAGTCTTCTTTGTTTCTCAATTTATTC
 ATC

>G965 Amino Acid Sequence (domain in AA coordinates: 423-486)

MGLATTTSSMSQDYHHHQGIFSFNNGFHRSSSTTHQEEVDESADVSGAQIPVYETAGMLS
 EMFAYPQGGGGGSGGEILDQSTKQLLEQQNRHNNNNSTLHMLLPNHQGFATDENTMQ
 PQQQQHTFWPSSSDHHQNRDMIGTVHVEGGKGLSLSLSSSLAAAKAEFYRSIYCAAVDG
 TSSSSNASAHHQFNQFKNLLLENNSSQHSHHHQVVGHFSSSSSPMASSSIGGIYTLRN
 SKYTKPAQELLEEFCSVGRGHFKNKLNRNNSNPNTTGGGGGGSSSSAGTANDSPPLSP
 ADRIEHQRRKVKLLSMLEEVDRRYNHYCEQMOMVNSFDQVMGYGAAPVYTTLAQKAMSR
 HFRCLKDAVAVQLKRSCELLGDKBAAGAASSGLTKGETPRLRLLEQSLRQRAFHHMGM
 EQEAWRPQRLPERSVNILRAWLFEHFLNPYPSPDADKHLARQTGLSRNQVSNWFINARV
 RLWKPMVEEMYQQEAKEREEAEEENENQQQORRQQTNMNDTKPNNNNNFVITAQTPT
 TMTSTHHENDSSFLSSVAAASHGGSDAFTVATCQQQDVSDFHVDGDGVNVIRFGTKQTDV
 SLTLGLRHSNIPDKNTSFSVRDFGDF*

>G1143 (54..677)

AAATAAGAATATAAACACTTTTGTCTGAAAAATTATCAAAGAAGAAGAAATAAATGGGTG
 GAGGAAGCAGATTTCAAGAACCAGTGAGGATGAGCCGTAGGAAACAAGTAACAAAAGAGA
 AGGAAGAAGATGAAAACCTTCAAACTCCAAATCTTGAAGCAGAGAGACGTAGAAGAGAGA
 AGCTTCATTGTGCGCTTATGGCTCTGCGATCTCATGTCCCCATTGTACCAACATGACTA
 AAGCAAGTATTGTTGAAGATGCGATTACTTACATAGGAGAGCTTCAAAACAATGTTAAGA
 ATCTCTTAGAGACATTTATGAAATGGAAGAAGCTCCTCCTGAGATTGATGAAGAACAAA
 CGGATCCAATGATAAAACCTGAAGTTGAAACTAGTGATCTTAACGAAGAGATGAAGAAAC
 TCGGAATCGAGGAGAATGTGCAATTGTGTAAGATTGGGGAGAGGAAGTTTTGTTTAAAGA
 TCATAACAGAGAAGAGAGATGGGATCTTTACTAAATTCATGGAGGTTATGAGATTTCTCG
 GATTCGAGATTATCGATATTAGTCTAACAACCTTCAAATGGAGCAATTCTTATTAGTGCCT
 CTGTTTCAGACACAGGAACCTCTGTGATGTTGAACAGACAAAAGATTTTCTTTTGAAGTTA
 TGAGAAGCAATCCATAAGTATTAAATTATATACATCTTGGAATTTCTTGATCTAATAACA
 TTTCCATTGGTTTTTATTACATTGTTGTTCCATTTTAAATATGATATGATTTCAGATGAAA
 AAGAGTTTGTGTTACAAGCCAATGA

>G1143 Amino Acid Sequence (domain in AA coordinates:33-82)
 MGGGSRFQEPVRMSRRKQVTKEKEEDENFKSPNLEAERRRRREKLHCRLMALRSHVPIVTN
 MTKASIVEDAITYIGELQNNVKNLLETFHMEEEAPPEIDEEQTDPMIKPEVETSDLNEEM
 KKLGIENVQLCKIGERKFWLKIITEKRDGIFTKFMEVMRFLGFEEIIDISLTSNGAILI
 SASVQTQELCDVEQTKDFLLEVMSNP*

>G1190 (209..2020)

TCCTGTCCCAAAACCAAGACTTGAGAGTGTGTCTTTAGAGAGAGATCTTCTCTCTTTT
 ATCTTACGACTCTCACTTCTTATCTCAAATCTACTTCAACTCTATTTCCAGTCTCCACAT
 TTTCCACAAATTTCAACTCTTGTCTCTTCTCCAAAGTAAAAACAAATCGTTGCAAG
 TGAGGTTTGGTTTTGGTGTATAGAATTATGAAGAGCGGGAAGCAATCTTCGCAACCTGA
 AAAGGGTACTTCCAGGATCTTGTCACTGACTGTCTGTTTATCGCATTTTGCGGTTTCTC
 CTTCTACCTCGGTGGTATATTTTGCTCTGAGAGAGACAAGATTGTAGCCAAGGATGTCAC
 AAGGACGACTACAAAGGCTGTAGCTTCCCTTAAAGAACCTACAGCTACTCCTATTCAAAT
 CAAATCCGTTTCTTTCCCGGAGTGCAGGTCAGAGTTCCAAGATTACACCCCGTGCACCGA
 TCCAAAGAGGTGGAAGAAGTATGGTGTCCATCGCTTAAGTTTCTTGGAGCGTCATTGTCTC
 TCCGGTATATGAAAAGAATGAGTGTGATTCCACCACAGACGGGTATAAACCGCCTAT
 AAGATGGCCCAAGAGCCGAGAACAGTGTGGTACAGGAACGTGCCTTATGATTGGATCAA
 TAAGCAAAGTCTAACCAGCATTTGGCTTAAGAAAGAAGGAGATAAGTTCCATTTCCCTGG
 TGGTGGTACCATGTTCCCTCGTGGAGTTAGTCACTATGTTGATTGATGCAAGATCTGAT
 TCCTGAAATGAAAGACGGAACAGTCAGGACCGCCATTGATACTGGCTGTGGGGTTGCGAG
 CTGGGGAGGCGATCTTTTGGACCGTGGGATACTATCACTCTCTTCTTGGCTCCAAGAGATAA
 CCATGAAGCTCAGGTTCAATTTGCTCTTGAACGTGGAATTCCTGCGATTCTCGGGATCAT
 CTCTACGCAACGTCTCCCTTTTCTTCAAATGCATTTGATATGGCTCATTGTTCAAGATG
 TCTTATTTCCCTGGACAGAATTTGGTGAATCTATTTACTTGAGATTACCGTATAGTTCCG
 ACCTGGAGGTTTTTGGGTTCTTTCTGGTCCACCTGTGAACTATAATAGACGATGGCGTGG
 ATGGAACACCAACCATGGAAGATCAGAAATCTGACTACAACAAGCTTCAGTCACTTCTAAC
 CTCCATGTGTTTCAAAAAGTACGCTCAAAAAGATGACATAGCCGTGTGGCAGAACTCTC
 AGACAAATCTTGCTATGACAAAATCGCTAAGAACATGGAAGCTTACCCTCCCAATGTGA
 CGACAGTATAGAACTGATTCTGCTTGGTACACTCCACTCCGTCTTGCCTGGTTGCCCGG
 GACACCTAAAGTCAAGAAGTCTGGTCTCGGATCAATCCCAAAATGGCCCCGAGAGGTTACA
 TGTCGCGCCCGAGAGAATCGGTGATGTTACCGGAGGGAGTGCAGAACAGTTTGAAACACGA
 TGATGGTAAATGGAAGAACAGAGTTAAGCATTACAAGAAAGTTTACCAGCTCTTGGGAC
 AGACAAGATAAGAAATGTTATGGATATGAACACTGTTTATGGAGGTTTCTCTGCGGCCCT
 CATTGAGGATCCCATTTGGGTCATGAACGTTGTATCATCGTACAGCGCAAATTCGCTTCC
 TGTGTCTTTGATCGCGGTCTCATCGGGACTTACCACGACTGGTGCAGAGCTTTCTCAAC
 GTATCCAAGACATATGATCTTCTTACCTCGACAGTCTTTTTACCTTGGAGAGTCAAG
 GTGTGAGATGAAGTACATTTTGCTAGAGATGGACAGGATCTTGCAGCCGAGTGGATATGT
 TATAATCCGAGAATCGAGTTATTTTATGACGCAATCACAACGTTAGCGAAAGGGATAAG
 GTGGAGTTGCCGAGAGAGGAGACTGAGTATGCAGTCAAAAGTGAGAAGATTCTGGTTTG
 CCAGAAAAAGCTATGGTTTTCGTCAAACCAAACCTCTTGATGAGACCACCTGTATCATAG
 TGTTTATCATCTCCTGTGATGCACACTACAGAGAGAAGGATCTAGTCCTTTGAGTCCAAG
 ATATAGCTCTATAACAAATCTCCTTTTTTTTGTCTCTTTAATTTCTTGGGTATTTACGG
 TATAGATTGATATTATATATTTTTTAATTATATTTTTTAATATATAGATATATTAGTATGT
 GGTTTAAACACTATTATTATCAAGGTCTTAAAGATTGCTTTGCAAGAGTTAAAAATGT
 TGGAGTAAGGACCTCTTGATTAATAAATTGACTGACGCAGCAA

>G1190 Amino Acid Sequence (domain in AA coordinates: entire protein)

MKSGKQSSQPEKGTSLRLSLTVLFIAFCGFSFYLGIGFCSEKIVAKDVTRTTTKAVAS
 PKEPTATPIQIKSVSFPECGSEFQDYTPCTDPKRWKYGVHRLSFLERHCPPVYEKNECL
 IPPPDGYKPIIRWPKSREQCWYRNPYDWINKQSNQHWLKKEGDKFHPGGGTMPFRGV
 SHYVDLMQDLIPMKDGTVRTAIDTGCVASWGGDLLDRGILSLSLAPRDNHEAQVQFAL
 ERGIPAILGIISTQRLFPFNSAFDMAHCSRCLIPWTEFGGIYLLIHRIVRPGFWVLSG
 PPVNYNRRWRGWNTTMDQKSDYNKLQSLLTSMCFKKYAQKDDIAVWQKLSKSCYDKIA
 KNMEAYPPKDDSI EPDSAWYTPLRPCVVAPTPKVKSGLSIPKWPERLHVAPERIGDV
 HGGANSLSKHDDGKWNKRVKHYKVLPA LGTDKIRNVMDMNTVYGGFSAALIEDPIWVMN
 VVSSYSANSLPVVDFDRGLIGTYHDWCEAFSTYPRTYDLLHLSLFTLESRCMKYILLE
 MDRILRPSGYVIIRESSYFMDAITTLAKGIRWSCRRETEYAVKSEKILVCQKKLWFSSN
 QTS*

>G1198 (230..1675)

TCTTTTCAAATTCCAATCATTTGATCAACTAATCAAGAATTAATTATAAGACTTTGCAAT
CTCTCTCCCTCTCCCTCTCCCTAGCTAGTTCTCTCTTGTGTTTCTTAACCTCGAGCTTCTC
TCAATAGTGATTATCATCTTTTTTCATCATTTCAAGATTTAATGTGTTTTGCAGAAAAGAG
ACTAATCAAGAAGAGATATCATCAATTGAAGCTGTTTTCTTGAGTAGAGATGGCGAACCA
TAGAATGAGCGAAGCTACAAACCATAACCACAATCATCATCTTCCTTATTCACTTATTCA
TGGTCTCAACAACAATCATCCATCTTCTGGTTTCATTAACCAAGATGGATCGTCCAGTTT
CGATTTTGGAGAGCTAGAAGAAGCAATTGTTCTGCAAGGTGTCAAGTATAGGAACGAGGA
AGCCAAGCCACCTTTATTAGGAGGAGGAGGAGGAGCTACGACTCTGGAGATGTTCCCTTC
GTGGCCAATCAGAACTACCAAACCTTCTCTACTGAGAGTTCCAAGTCAGGAGGAGAGAG
CAGCGATTGAGGATCGGCTAATTTCTCCGGCAAAGCTGAAAGTCAACAACCGGAGTCTCC
TATGAGTAGCAAAACATCATCTCATGCTTCAACCTCATCATAATAACATGGCAAACCTCAAG
TTCAACATCTGGACTTCTTCCACTTCTCGAAGTTTAGCTCCTCTAAACCTTCGGAAGA
TAAGAGGAAGGCTACAACCTCAGGCAACAGCTTGATGCTAAGACGTTGAGACGTTTGGC
CCAAAATAGAGAAGCTGCTCGCAAAAGCCGTCTTAGGAAAAAGCGTATGTGCAACAGCT
AGAATCAAGTAGGATAAAGCTTTCCCAATTGGAGCAAGAACTTCAGCGAGCTCGTTCTCA
GGGGCTGTTTATGGGTGGTTGTGGACCACCAGGACCTAACATCACTTCCGGAGCTGCAAT
ATTTGACATGGAATATGGGAGATGGCTAGAGGATGATAACCGGCATATGTGCGAGATTCTG
AACCGGTCTTCAGGCTCATTTATCTGACAATGATTAAAGGTTGATCGTTGACGGTTACAT
TGCTCATTTTGTGATGAGATATCCGATTAAAAGCCGTGGCAGCGAAAGCCGATGTTTTTCA
CCTCATCATTTGGGACATGGATGTCCCCAGCCGAACGTTGTTTTATTGGATGGCTGGTTT
CCGTCCATCCGACCTAATCAAGATATTGGTGTGCGAAATGGATCTATTGACGGAGCAACA
ACTGATGGGAATATATAGCCTACAACACTCGTTCGCAACAAGCAGAGGAGGCTCTCTCGCA
AGGCCTCGAACAACCTTCAGCAATCTCTCATCGATACTCTCGCCGCACTCTCCAGTCATTGA
CGGAATGCAACAAATGGCTGTGCTCTCGGAAAGATCTCTAATCTCGAAGGCTTTATCCG
CCAGGCTGATAACTTGAGGCAGCAGACCGTTACCAGCTGAGGCGGATCTTGACCGTCCG
ACAAGCTGCACGGTGTTCCTAGTCATCGGAGAGTACTATGGACGGCTCAGAGCTCTTAG
CTCCCTTTGGTTGTACGCCCCAGAGAGACTGATGAGTGATGAAACCTCTTGTCAAAC
GACGACGGATTTCAGATTGTTTCAGTCATCTCGGAACCACTTCTCCAATTTCTGAATGGA
ATGAACTTTGTATAACTAAAAGGCCAAGTTTCATTGTCTGTGCGTAATTTACCTATTTT
CTTTAAAGTTGTACTAGAGAAAAGATAGGATCTTCTCTCG

>G1198 Amino Acid Sequence (domain in AA coordinates: 173-223)
MANHRMSEATNHNHNNHLPYSLIHGLNHNHPSSGFINQDSSSFDGFELEEAIVLQGVKY
RNEEAKPPLLGGGGGATTLEMFPSWPIRTHQTLPTESSKSGGESSDSGSANFSGKAESQQ
PESPMSKHHMLQPHHNNMANSSSTGLPSTSRTLAPPKPSKEDKRKATTSKGQLDAKTL
RRLAQNRREARKSRLRKKAYVQQLSSRIKLSQLEQELQARSQGLFMGGCGPPGPNITS
GAAIFDMEYGRWLEDDNRHMSEIRTLQAHLSNDLRLIVDGYIAHFEIIFRLKAVAAGA
DVFHLLIGTWMSPAERCFIWMAGFRPSDLIKILVSQMDLLTEQQLMGIYSLQHSQQABE
ALSQGLEQLQQLIDTLAASPVIDGMQMAVALGKISNLEGFIRQADNLRQQTVHQLRRI
LTVRQAARCFIVIGEYYGRLRALSSLWLSRPRETLSDETSCQTTTDLQIVQSSRNHFSN
F*

>G1226 (212..1159)

CTGCATTTATTAAGAACAGTTTAGAAAAGTGTCAACCCCTAAAGGAATGTTTTTAGTTTAG
AGGAAAGAGAGAGAAGAAGAGCAGCAGCAGAAGTTGTTAATTTGAAGACTATTTGAGGA
AAGACACCTATATCTAAATACTCAAAGTTACAAAAATATTACTTCAGAAAAAGTTCCAT
TAGAGAGACTCATAAAGCTTCTCATCTAATTATGAGTGGATTGATGAGTTTGGTGAATT
AGAAGACCAATTTGGTCAGATTTTCAGACACTACTATGGAAGAGAAGATACCATTTCTGCA
AATGCTTCAATGCATAGAACACCCCTTTTACAACAACAGAACCAATCAGTTTCTCCAATC
ACTTCTCCAGATCCAAACCTTAGAATCAAAGAGCTGTCTCACCCCTTGAAACAAACATCAA
AAGAGATCCGGGTCAAACAGATGACCCGGAAGGATCCAAGAACAGAAAACGGAGCAGT
AACGGTCAAAGAAAAAGAAAACGGAAACGTACAAGAGCTCAAAGAAACAAAGACGAAGT
TGAAAACCAAAGGATGACTCACATTGCCGTGCAACGTAATCGAAGACGACAAATGAACGA
ACACTTAACTCTCTCGATCTCTCATGCTCCTTCTGTTTCTTCAACGGGGTGACCAAGC
TTCGATTGTAGGAGGGGCAATAGATTTCATCAAGGAAGTAGAGCAACTCTTGCAATCTCT
AGAAGCTGAGAAACGAAAGGATGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT
ATCTTCGTCTCTTGCATGCACTAACTCTTCTATTCTTAGCGTGTCTACGACGTGCGAAAA
TGGATTTACGCGGAGATTGCGCGGTGGAGATACGACAGAAGTGGAGGCTACGGTGATACA

GAACCATGTGAGCTTAAAAGTTCCGGTGTAAAGAGAGGAAAACGACAGATCTTAAAAGCTAT
TGTCTCGATTGAAGAACTAAAGCTTTCGATTCTACATCTCACTATCTCTTCTTCCCTTGA
CTTTGTGCATCTACTCTTTCAATCTCAAGATGGAAGATGGTTGTAAATTAGGATCAGCAGA
TGAGATAGCGACAGCCGTTTCATCAGATCTTCGAGCAAATCAACGGTGAAGTCATGTGGTC
AAATCTTAGTCGAACCTAGTTGACTTTTGACTCCTAGTAACGTGTGTAAACTTTAGGTTA
CAAAGAAAAGGGACGTGATATAAATAAGAAAAACCAAAGAGGTGAAATTTTGGGAGTTT
AATTATTATCTTATACTTTTTGGATTTTAGATTAGTAGCAAACCTCGCAGTGTCTACGAT
GACATTATTATTGGTCACATGAAGGTTTAGGTTAAAAA

>G1226 Amino Acid Sequence (domain in AA coordinates:115-174)
MSGLMSFGELEDQFGQISDTTMEEKIPFLQMLQCIEHPFTTTEPNQFLQSLLQIQTTLESK
SCLTLETNIKRDPGQTDDEKDPRTENGAVTVKEKRKRKRTRAPKNKDEVENQRMTHIAV
ERNRRRQMNEHLNLSRLMPPSFLQRGDQASIVGGAIDFIKELEQLLSLEAEKRKDGTD
ETPKTASCSSSSSLACTNSSISSVSTTSENGFTARFGGGDTTEVEATVIQNHVSLKVRCK
RGKRQILKAIVSIEELKLAILHLTISSSFDVFIYSFNLKMGEDGCKLGSADDEIATAVHQIF
EQINGEVMWSNLSRT*

>G1451 (124..2559)

TTTGTACTTCCGGAGCTAAAGAGTTATAGCTACTGTAGTAGCTGGAAGTGAAGAAGATTT
TTTAATAGATTGTACGGAATAATAGGGTTTTCAAAGTTTGGTTTCTTGAAGTTGAATTA
GACATGAAGCTGTCAACATCTGGATTGGGTCAACAGGGTCATGAAGGAGAGAAGTGTCTG
AATTCTGAGCTATGGCATGCTTGTGCTGGACCATTAGTCTCTCTTCCATCATCTGGTAGT
CGAGTTGTTTTACTTTCCACAGGGTCACAGTGAACAGGTAGCTGCTACAATAAAGGAA
GTTGATGGTCACATACCCAATTACCCAAGCCTACCACCACAATTGATATGCCAGCTCCAT
AATGTTACAATGCATGCAGATGTTGAGACGGATGAAGTCTATGCTCAAATGACACTTCAA
CCATTGACACCGGAGGAGCAGAAAGGAAACATTTGTACCGATTGAGTTGGGGATACCGAGT
AAGCAACCTAGTAATTATTTTTGTAAAGTCTCTACAGCTAGTGATACCAGTACACATGGA
GGGTTTTCTGTTCTAGACGTGCTGCTGAGAAAGTGTTCCTCCATTGGATTACACACTG
CAGCCACCAGCTCAAGAACTGATTGCAAGGGATCTCCATGATGTTGAATGGAAGTTTAGG
CATATCTTTTCGGGGACAGCCCAAACGGCATCTCCTAACTACTGGATGGAGTGTCTTTGTC
AGTGCCAAGCGACTAGTAGCTGGAGATTCTGTCAATTTTCATCAGGAATGAAAAGAATCAA
CTCTTTTTTGGGAATTCGTATGCCACTCGGCCGAGACTATTGTACCATCATCTGTTTTA
TCTAGTGATAGCATGCATATTGGACTCCTTGCTGCTGCTGCACATGCTTCTGCAACTAAT
AGCTGTTTCACTGTTTTCTTTCATCCAAGGGCTAGCCAATCTGAGTTTGTGATACAATT
TCCAAGTACATTAAAGCCGTTTTTTCACACGCGTATTTTCAGTTGGGATGCGTTTTCGCATG
CTCTTCGAGACAGAAGAGTCGAGTGTCCGCGAGTACATGGGTACTATAACTGGTATTAGT
GATCTAGATTCTGTTTCGTTGGCCAACTCTCATTGGCGATCTGTGAAGGTTGGTTGGGAT
GAATCGACTGCAGGGGAGAGACAGCCAAGGGTTCTTTATGGGAGATTGAGCCTCTGACT
ACCTTTCCTATGTATCCATCTCTTTTTCTCTCAGACTAAAACGTCCATGGCATGCTGGC
ACATCATCTTTGCCTGATGGAAGGGGTGATTGGGAAGTGGTCTAACATGGCTAAGAGGG
GGAGGTGGAGAGCAGCAAGGTTTGCTTCTCTAAATTATCCATCTGTTGGTTTGTTCCTCA
TGGATGCAACAAAGGCTGGATCTCAGTCAAATGGGGACTGATAATAATCAGCAATACCAA
GCAATGTTAGCTGCTGGGTTGCAGAACATCGGCGGTGGAGATCCTTTAAGACAGCAGTTT
GTACAGCTGCAAGAGCCTCACCACCAATATCTTCAACAATCAGCTTCCCATAATTCTGAT
TTGATGCTTCAGCAGCAACAGCAGCAACAAGCGTCACGCCATCTCATGCATGCTCAAACA
CAGATTATGAGTGAGAATCTCCGAGCAGAATATGCGACAAGAAGTTAGTAACCAACCA
GCTGGACAGCAGCAACAGCTACAGCAACCGGACCAAAATGCATATCTTAATGCTTTCAAA
ATGCAAAATGGCCATCTTCAACAGTGGCAGCAGCAATCAGAGATGCCATCTCCCTCGTTC
ATGAAGTCAGATTTTACTGACTCAAGCAACAAATTTGCAACAACTGCTAGTCCGGCTTCT
GGAGATGGCAATCTTTTGAATTTTCTATAACCGGTCAGTCTGTACTCCCTGAGCAGTTA
ACAACAGAGGGCTGGTCTCCAAAGCATCCAACACTTTTCTGAACCGTTGTCACTTCCA
CAAGCCTATCTGGGAAGAGTCTTGCTCTAGAACCCGGAATCCGAGAAATCCCTCTCTT
TTGCGGTGTTGATCCCGACTCTGGACTCTTCTCCCGAGTACGGTCCCGCTTTGCTTCT
TCATCAGGAGATGCTGAAGCTTCCCTATGTCACTAACAGATTCAGGATTTTCAGAATTCC
TTATATAGCTGCATGCAAGACAACTCATGAGTTATTGCATGGAGCTGGACAGATTAAC
TCGTCCAACCAACCAAGAACCTTTGTAAAGGTTTATAAATCTGGTTCCGGTTGGGCGTTCA
TTAGACATCTCCGATTGAGCAGCTACCACGAGCTGCGAGAAGAGTTAGGGAAGATGTTT
GCTATCGAAGGGTTGTTGGAAGACCCCTTAGATCAGGCTGGCAGCTTGTATTGTTGAC
AAGGAAAATGATATCTTCTCTTGGTGATGACCCATGGGAGTCATTTGTGAATAACGTT

TGGTACATAAAGATACTATCACCAGAAGATGTGCATCAAATGGGAGATCATGGAGAAGGC
AGTGGTGGGTTATTCCCGCAAAACCCGACCCATCTCTAGAAGCTGCTTCGGTGTAGTCT
CATCATGCTACAACGCGGAGCCCTTTGTTTCCCATTTGAAGTCGTTTCCACTCATCTTT
ATATGCCATTCTGTCGCATCTCTCTCGTTTGGACGTTTTAGAAAGAAACATAATCATAT
TTGTGAGTATGGGTCTGAAACTTTAGGACGTACTTTAGCTTGTATTAGACAGACACTCT
CGTCATAAACATAAGAACCTTTATGTAGCTGTCTCAGGGTAACTAACTTTTCTAG
>G1451 Amino Acid Sequence (domain in AA coordinates: 22-357)
MKLSTSGLGQQGHEGKCLNSELWHACAGPLVSLPSSGSRVYFPQGHSEQVAATTNKEV
DGHIPNYPSPQLICQLHNVTMHADVETDEVYAQMTLQPLTPPEQKETFPVPIELGIPSK
QPSNYFCKTLTASDTSTHGGFVSPRRAAEKVFPPLDYTLQPPAQELIARDLHDVEWKFRH
IFRGQPKRHLTTGWSVVFSAKRLVAGDSVIFIRNEKNQLFLGIRHATRPQTIVPSSVLS
SDSMHIGLLAAAHASATNSCFTVFFHPRASQSEFVIQLSKYIKAVFHTRISVGMFRML
FETEESVRRYMGITITGISDLSVRWPNHRSVKVWDESTAGERQPRVSLWEIEPLTT
FPMYPSLFLRLKRPWHAGTSSLPDGRDLGSLTTLRGGGGEQQGLPLNYPVSVGLFPW
MQQRDLDSQMGTDDNNQOYQAMLAAGLQNIIGGDPLRQGFVQLQEPHHQYLOQSASHNSDL
MLQQQQQQQASRHLMAHTQIMSENLPQNMQRQVSNQPAQQQQQLQPPDQNAYLNAFKM
QNGHLLQQWQQQSEMPSPSFMKSDFTDSSNKFATTASPASGDNLLNFSITGQSVLPEQLT
TEGWSPKASNTFSEPLSLPQAYPGKSLALEPGNPQNPSLFGVDPDSGLFLPSTVPRFASS
SGDAEASPMSLTDSGFQNSLYSCMQDTTHELLHGAGQINSSNQTKNFVKVYKSGSVGRSL
DISRFSSYHELREELGKMFIEGLLEDPLRSGWQLVFVDKENDILLGLDDPWESFVNNVW
YIKILSPEDVHQMGDHGEGSGGLFPQNPTHL*
>G1478 (1..354)
ATGTGTAGAGGGTTTGAGAAAGAAGAGAGAAGAAGCGACAATGGAGGATGCCAAAGA
CTATGCACGGAGAGTCACAAAGCTCCGGTAAGCTGTGAGCTTTGCGGCGAGAACGCCACC
GTGTATTGTGAGGCAGACGCAGCTTTCTTTGTAGGAAATGCGATCGATGGGTCCATTCT
GCTAATTTTCTAGCTCGGAGACATCTCCGGCGCGTGATCTGCACGACCTGTGCGAAGCTA
ACTCGTCGATGTCTTGTGCGGTGATAATTTTAAATGTTGTTTTACCGGAGATAAGGATGATA
GCAAGGATTGAAGAACATAGTAGTGATCACAAATTCCTTTGTGTTTCTCTGA
>G1478 Amino Acid Sequence (domain in aa coordinates: 32-76)
MCRGFEKEEERRSDNGCQRLCTESHKAPVSCELCGENATVYCEADAAFLCRKCDRWVHS
ANFLARRHLRRVICITCRKLTRRCLVGDNFNVVLPEIRMIARIEEHSSDHKIPFVFL*
>G1496 (116..1123)
AAACCCACCAAATAACTCAGAGCTTTTTTGCAATTTTTTCCCATTTCTATTTTGT
ACTTTTGGTCTCACTTTAAAGATCATAAGTTGAAAGATTTCTGCAGAGAACAATATGTT
GGAAGGTCTTGTCTCTCAAGAAAGCTTGTCTTAAACTCTATGGACATGTCTGTACTTGA
AAGGCTTAAATGGGTACAACAGCAACAACAGCAACTGCAACAAGTTGTGTCCCATAGCAG
TAATAATTCACCTGAACCTTCTCAGATACTTCAGTTCCATGGAAGCAACAATGATGAGTT
GTTGGAGAGTAGTTTTCAGCCAATTTCAAATGCTTGGATCTGGTTTTGGACCAAACATATAA
CATGGGTTTTTGGTCTCCACATGAATCCATTTCAAGAACAAGTAGCTGCCATATGGAACC
TGTGGATACAATGGAGGTTTTGTTGAAGACCGGTGAAGAAACCAGAGCCGTTGCCTTGAA
GAACAAGAGAAAACCAGAGGTTAAGACAAGGAAGAGCAAAAGACAGAGAAGAAGATCAA
AGTAGAGGCTGAGACAGAGTCAAGCATGAAAGGAAAATCAAACATGGGAAACACTGAAGC
ATCTTCAGACACTTCAAAGGAGACATCGAAAGGAGCTTCAGAGAATCAGAAATTAGATTA
TATCCACGTGAGAGCTCGTCGAGGCCAAGCCACTGACAGACACAGCTTAGCAGAAAGGGC
GAGAAGAGAAAAGATCAGCAAGAAAATGAAATATCTGCAAGATATTGTGCCTGGATGCAA
TAAGGTCACAGGAAAAGCTGGTATGCTTGATGAGATCATCAATTATGTTCAATGTCTCCA
AAGACAAGTCGAGTCTCTGTGTCGATGAAACTTGTCTTGAACCGGAACCTAGAGCTTGC
CGTGAAGATGTATCCGTAAAACAGGCTTACTTTTCAAATGTAGTTGCTTCAAAGCAATC
AATAATGGTTGATGTGCCATTGTTTCCGTTAGACCAGCAAGGATCTCTAGATTGTCTGC
GATAAACCCGAACCAACGACATCTATCGAAGCTCCATCTGGAAGCTGGGAAACCTAATC
ACAGAGTCTCTACAACACATCTAGCCTCGGTTTTTCACTACTAAGCAAGATTCAATGAAAC
AACATGGTTGACATCAATCAATCATCAAAATCAGAAGCAAATTCATTACATTTGCTCAT
CAAAGTAGTAATTTGAAATTTGGTTAATGCATTATCCTTTGATCCTTGTCTTCTGATAT
TTAAACCAGAAGAACTGGAGATAGCAATCCAATGATCTTGTACCA
>G1496 Amino Acid Sequence (domain in AA coordinates: 184-248)
MLEGLVSQESLSLNSMDSVLERLKWVQQQQQLQVVS HSSNNSPELLQILQFHGSNND
ELLESSFSQFQMLGSGFGPNYNMFGPPHESISRTSSCHMEPVDMEVLLKTGEETRAVA

LKNKRKPEVKTREEQKTEKKIKVEAETESSMKGKSNMGNTASSDTSKETS KGASENQKL
DYIHVRARRGQATDRHSLAERARREKISKMKYLQDIVPGCNKVTGKAGMLDEIINYVQC
LQRQVEFLSMKLAVLNPELELAVEDVSVKQAYFTNVVASKQSIMVDVPLFPLDQQGSDDL
SAINPNQTTTSIEAPSGSWETQSQSLYNTSSLGFHY*

>G1526 (1..3090)

ATGGGAACGAAAGTCTCAGACGATCTTGTTTCCACCGTCAGATCAGTCGTGGGTTCGGAT
TACTCAGATATGGATATAATCAGGGCTTTACACATGGCGAATCATGATCCAACGGCTGCT
ATCAATATAATCTTCGACACTCCAAGTTTCGCCAAACCTGATGTAGCCACTCTACCCCG
AGCGGCTCTAATGGAGGGAAGCGAGTTGATAGTGGATTAAAGGGCTGTACTTTTGGTGAC
AGCGGAAGTGTGGAGCGAATCATCGCGTGGAGGAAGAAAATGAGAGTGTAAATGGTGGA
GGAGAAGAGAGTGTTCAGGGAATGAGTGGTGGTTTGTGGTTGTTCTGAATTGGCTGGG
TTATCGACATGTAAAGGAAGGAAATGAAGTCTGGTGATGAATTGGTGTTCACGTTTCCG
CATAGTAAAGGATTAAAGCCTGAGACTACGCCTGGGAAGCGCGGTTTGGGCGGGGAAGG
CCAGCTTTGCGTGGTGCTTCTGATATCGTTAGGTTCTCTACAAAGGATTTCAGGAGAGATT
GGTAGAATACCAAACGAGTGGGCTCGGTGTCTTCTACCACTTGTGAGAGACAAGAAAATT
AGGATAGAAGGCAGTTGCAAGTCGGCGCCTGAAGCTTTGAGCATCATGGATACAATTCTT
CTGTCTGTAAAGCGTGTACATTAATAGTTCCATGTTTCAAAGCATAGTGGACTTCATTT
AAGACAGCTAGTAATACGGCAGAGGAATCAATGTTCCATCCTCTCCCAAATCTCTTCGG
TTACTCGGTTTGATCCCTTTAAGAAGGCAGAGTTTACTCCAGAGGATTTTACTCTAAG
AAGCGACCTTTGAGTTCCAAGGATGGTTCTGCTATTCTTCTTCTGCTTCAATTAAAC
AAGGTCAAGAATATGAATCAAGATGCAAACGGAGATGAAAATGAGCAGTGTATCAGCGAT
GGTGATCTTGATAACATTGTTGGTGTGGGGACAGTTCGGATTAAAGGAAATGGAACT
CCACATACACTTCTGTGTGAGCTTCGTCCATACCAAAGCAGGCACCTTCATTGGATGACC
CAACTGGAGAAAGGAAATTGCACTGATGAGGCAGCAACAATGCTTCACCCGTGTTGGGAA
GCATAGTGTAGCAGACAAGAGGGAAGTGGTTGTCTACCTGAATTCTTTTACTGGTGAT
GCTACAATACACTTCCCTAGCACACTTCAAATGGCAAGAGGAGGAATATTAGCAGACGCA
ATGGGTCTTGGAAGAGACTGTAATGACCATATCCCTTTTGCTTGCCCATCTTGGAAGCT
GCATCAACTGGGTTTCTATGCCCCAACTATGAAGGAGACAAAGTGATCAGCAGTTCTGTA
GATGATCTCACTAGTCCCCCGGTGAAGGCAACCAAATTTCTAGGCTTTGATAAGAGGCTT
CTTGAACAAAAAGTGTACTTCAAATGGTGGTAACCTGATTGTATGTCCGATGACACTT
TTAGGACAGTGGGAAGACAGAGATTGAAATGCATGCAAAGCCTGGGTCTCTATCTGTCTAT
GTTCACTATGGGCAAAGCAGGCCGAAGGATGCAAACCTTCTTCCAGAGTGATGTGGTA
ATCACCATATGGAGTTCTAACATCCGAATTCTCGCAAGAGAACTCAGCAGACCATGAA
GGAAATTTATGCAGTTTCGATGGTTTAGGATTGTTCTTGACGAGGCACATACCATCAAAAC
TCAAAAAGCCAAATTTCTTGGCTGCTGCAGCTCTGGTTGCTGATAGGCGTTGGTGTCTT
ACGGGTACTCTTATTCAGAACAACTCTGGAGGATTATACAGCCTTCTACGGTTTTTGAGG
ATTGAACCATGGGGAACCTGGGCATGGTGGAAATAAAGTTGTCCAAAAGCCATTTGAAGAG
GGTGATGAGAGAGGGTTAAAGCTAGTGCACTCTATCTTAAACCTATCATGCTTAGGAGA
ACAAAGTCTAGCACAGACCGAGAAGGAAGGCCGATTCTTGTTCTACCCCTGCTGATGCA
CGGGTCAATTTACTGTGAACCTTTCGGAGTCTGAGAGGGATTCTACGACGCGCTATTTAAA
AGATCCAAGGTCAAATTTGATCAATTTGTTGAACAAGGCAAAGTTCTTCATAACTATGCT
TCGATCCTGGAACCTGCTTTGCGTCTTCGACAATGTTGTGATCACCCTTTTGTAGTAATG
AGTCGAGGGGATACAGCGGAATACTCTGATCTGAATAAGCTTTCTAAACGTTTCCTTAGT
GGAAAGTCTTCTGGCTTAGAAAAGGGAAGGAAAAGATGTACCGTCAGAGGCTTTTGTTCAG
GAGGTGGTAGAGGAACGCGCAAAGGAGAGCAAGGAGAGTGCCAATATGCCTTGAAGCA
CTTGAGGATGCTGTATTAAACGCCATGTGCTCATAGATTATGTGCTGAGTGTCTCTTGGCA
AGTTGGAGAAATTTACTTCTGGGTTATGTCTGTGTAGGAACACTGTAAGCAAACAA
GAACATCATCAGCACCAACCGAAAGTAGATTCCAGGTTGACGTGGAAAAGAAATGGGTG
GAATCATCGAAAATCACTGCTCTTCTGGAAGAGCTTGAAGGTCTTCGTTCTTCAGGCTCT
AAGAGCATTCTCTTTAGCCGATGGACCGCTTTCTCGATCTCTCTCAAATTTCCCTCTCT
CGGAATAACTTTTTCATTGTCCTGCTTGTATGGCACGCTAAGTCAGCAGCAACGAGAGAAG
GTCTTTAAAGAATTTTCCGAAGATGGCAGTATCCTGGTACTGTTGATGTCTCTAAAAGCT
GGTGGCGTTGGGATAAATCTAACAGCTGCGTCCAATGCTTTGTGATGGATCCATGGTGG
AACCAGCGGTAGAGGAACAAGCTGTTATGCGTATTTCATCGTATAGGGCAAACCTAAGGAA
GTCAAATCAGAAGATTTCATCGTTAAGGGAACGGTTGAAGAGAGAATGGAGGCGGTTTCAG
GCGAGGAAGCAGAGAATGATCTCTGGGGCTTTAACCGATCAAGAAGTACGAAGTGCACGT
ATAGAGGAACCTCAAGATGTTATTTACCTGA

>G1526 Amino Acid Sequence (domain in AA coordinates: 493-620, 864-1006)
MGTKVSDDLVSTVRSVVGSDYSDMDIIRALHMANHDPTAAINIIFDTPSFAKPDVATPTP
SGSNGGKRVDSGLKGCTFGDSGSGVGNHRVEEENESVNGGGEESVSGNEWWFVGCSELG
LSTCKGRKLKSGDELVFTFPHSKGLKPETTPGKRGFGRGRPALRGASDIVRFSTKDSGEI
GRIPNEWARCLLPVLRDKKIRIEGSCSAPEALSIMDTILLSVSVYINSSMFQKHSATSF
KTASNTAEESMFHPLPNLFRLLGLIPFKAEFTPEDFYSKRPLSSKDGSAPIPTSLQLN
KVKNMNQDANGDENEQCISDGLDNIVGVGDSSGLKEMETPHTLLCELRPYQKQALHWM
QLEKGNCTDEAATMLHPCWEAYCLADKRELVVYLNSFTGDATIHFPSTLQMARGGILADA
MGLGKTVMTISLLLAHSWKAASTGFLCPNYEGDKVISSSVDDLTPSPVKATKFLGFDKRL
LEQKSVLQNGGNLIVCPMTLLGQWKTEIEMHAKPGSLSVYVHYGQSRPKDAKLSSQSDVV
ITTYGVLTSEFSGQENSADHEGIYAVRWFRIVLDEAHTIKNSKSQISLAAAALVADRRWCL
TGTPIQNLEDLYSLRFLRIEPWGTWAWWNKLQKPFEEGDERGLKLVQSILKPIMLRR
TKSSTDREGRPIVLPPADARVIYCELSERDFYDALFKRSKVKFDQFVEQGVHLHNYA
SILELLRLRQCCDHPFLVMSRGDTAEYSDLNKLSKRFLSGKSSGLEREGKDVPEAFVQ
EVVEELRKGEQGECPICLEALEDVLTCAHRLCRECLLASWRNSTSGLCPVCRNTVSKQ
ELITAPTESRFQVDVEKNWVESKITALLLEEGLRSSGSKSILFSQWTAFLDLLQIPLS
RNNFSFVRLDGTLSQQQREKVLKEFSEDGSILVLLMSLKAGGVGINLTAASNAFVMDPWW
NPAAVEEQAVMRIHRIGQTKVKIRRFIVKGTVEERMEAVQARKQRMISGALTDQEVRSAR
IEELKMLFT*

>G1543 (1..828)
ATGATAAACTACTATTTACGTACATATGCACATACACATATAAACTATATGCTCTATAT
CATATGGATTACGCATGCGTGTGTATGTATAAATATAAAGGCATCGTCACGCTTCAAGTT
TGCTCTCTTTTATATTAACCTGAGAGTTTTCTCTCAAACCTTTACCTTTTCTTCTCGATC
CTAGCTCTTAAGAACCCTAATAATTCATTGATCAAAATAATGGCGATTTTGCCGGAAC
TCTTCAAACCTTGGATCTTACTATCTCCGTTCCAGGCTTCTCTTCATCCCTCTCTCCGAT
GAAGGAAGTGGCGGAGGAAGAGACCAGCTAAGGCTAGACATGAATCGGTTACCGTCGTCT
GAAGACGGAGACGATGAAGAATTCAGTCACGATGATGGCTCTGCTCTCCGCGAAAGAAA
CTCCGCTCTAACCAGAGAACAGTCACGCTCTTCTGAAGATAGTTTCAGACAGAATCATACC
CTTAATCCCAAACAAAAGGAAGTACTTGCCAAGCATTTGATGCTACGGCCAAGACAAATT
GAAGTTTGGTTTCAAACCGTAGAGCAAGGAGCAAAATTGAAGCAAACCGAGATGGAATGC
GAGTATCTCAAAGGTGGTTTGGTTTCATTAACGGAAGAAAACACAGGCTCCATAGAGAA
GTAGAAGAGCTTAGAGCCATAAAGGTTGGCCCAACAACGGTGAACCTCTGCCTCGAGCCTT
ACTATGTGTCTCGCTGCGAGCGAGTTACCCCTGCGCGAGCCCTTCGAGGGCGGTGGTG
CCGTTTCCGGCTAAGAAAACGTTTCCGCGCAAGAGCGTGATCGTTGA

>G1543 Amino Acid Sequence (domain in AA coordinates: 135-195)
MIKLLFTYICTYTYKLYALYHMDYACVCMYKYKGIIVTLQVCLFYIKLRVFLSNFTFSSSI
LALKNPNNSLIKIMAILPENSSNLDLTISVPGFSSSPLSDESGSGGRDQLRLDMNRLPSS
EDGDDEEFSHDDGSAPPRKKLRLTREQSRLLDSFRQNHITLNPQKEVLAKHMLRPRQI
EVWFQNRARRSKLKQTEMECEYLKRWFGSLTEENHRLHREVEELRAIKVGPTTVNSASSL
TMCPRCERVTPAASPSRAVVPVPAKKTFFPPQERDR*

>G162 (101..619)
AGACATACAACACCAAATCTTCTTCTTCAACACATATTCACCTTTCACAGCAAAAAA
ACGAGAGGTTCTCTCTTATTCGTACCGTTTTAGCAAACAAATGGGTCCGAGAAAGATCAA
GATGGAGATGGTTCAAGGACATGAACACACGACAGGTTACCTTTTCAAACGGAGGACTGG
TTTGTTCAGAAGGCGAGCGAGTTAGCCACGCTCTGCAACGCTGAGTTGGGCATCGTTGT
CTTTTCAACAGGAGGCAAGCCTTCTCCTACGGGAAACCGAATCTTGATTCTGTTGCAGA
GCGATTATGAGAGAATATGATGATTGAGACAGTGGCGATGAAGAAAAAAGTGTAATTA
CAGGCTAAACTGAAGAGGCTGAGTGAACGCTCTCGATTGTCTCAACCAAGAGGTTGAAGC
TGAGAAGGAACGAGGCGAGAAGAGTCAGGAGAAGCTTGAATCTGCTGGGGATGAGAGATT
CAAGGAGTCCATTGAGACGCTTACCCTCGATGAACCTCAATGAATACAAAGATAGGCTTCA
GACAGTCCATGGTAGGATTGAAGGTCAAGTCAATCACTTGACGGCTTCGTCTTGCCTCAT
GCTTCTCTCCAGAAAATAGCTAGACCGACTTGTAGAGTTACATTCTATTTTTGTATCA
GCCTACAGAACTTACCAACACATGAAAGTTATTGCTGGTGTAGAATTTTCTGTCTATCTAT
GGGGTGTGACTTTCTATTGACATCAAATGAAAATGTACCTGGAAATTTGTCTGTATTAA
TCTCAAGTGTACTTGTAAACTTGATCAGCTTTTTTCGCAAAAAA

>G162 Amino Acid Sequence (domain in AA coordinates: 2-57)
MGRRKIKMEMVQDMNTRQVTFSKRRTGLFKKASELATLCNAELGIVVFSPPGKPFSSYKPK

NLDSVAERFMREYDDSDSGDEEKSGNYRPKLKRLSERLDLLNQVEAEKERGEKSQEKLE
SAGDERFKESIETLTLDLNEYKDRLQTVHGRIEQVNHQLQASSCLMLLSRK*

>G1640 (168..1196)

TTCCGCAGATCCTTCCTCTATATAAGGAAGTTCATTTTCATTTGGAGAGGTTTCGCTGACA
AGCTGCTCTAGCTTATCTGGTACCGTCGACCTCTCACTCAAGGGTCCAAAAGTGTCTTCT
CTTTTTTCAGTTTCTCTTTCTCTTTTGGACAGAAGAGACCGAGAAGCAATGGGAAGGGCTC
CGTGTGTGAGAAAATCGGGTTGAAGAGAGGGAGATGGACAGCCGAGGAAGATGAGATCC
TCACCAAGTATATTCAGACCAATGGTGAAGGTTCTTGGCGATCTTTGCCTAAGAAAGCTG
GATTGTTGAGATGTGGAAGAGCTGTAGACTAAGGTGGATAAACTACTTAAGAAGAGACT
TAAAAAGAGGAAATATTACTTCCGACGAAGAAGAAATAATCGTCAAGTTGCATTCCCTTC
TCGGCAACAGATGGTCACTTATTGCAACACATCTACCAGGAAGAACAGACAACGAAATTA
AAAATATTGGAACACATCTCAGCCGCAAAATCTATGCCTTCACTGCCGTTTCCGGAG
ATGGACACAATCTACTCGTCAACGATGTAGTCTTGAAGAAATCTTGTTTCATCGTCTTCTG
GAGCCAAGAACAATAACAAGACCAAGAAGAAGAAGGGAAGGACTAGTAGGTCAATCCA
TGAAGAAACACAAGCAAAATGGTGACGGCCTCACAATGTTTCTCACAACCTAAGGAGCTAG
AGAGTGATTTTCAGTGAGGGAGGGCAAAATGCTAATTTTGAAGGAGAGTCTTTGGGGCCTT
ATGAGTGGTTGGATGGTGAAGCTAGAACGGCTCTTGAGTAGTTGTGTCTGGGAATGCACTA
GTGAAGAGGCTGTGATTGGAGTAAATGATGAAAAGGTGTGTGAGAGTGGGGACAATAGTA
GTTGTTGTGTTAATTTGTTTGAAGAAGAACAAGGAAGCGAGACAAAGATTGGTCACGTAG
GAATCACAGAGGTTGATCATGATATGACGGTGGAAAGAGAAAGAGAGGGAAGTTTTTTAA
GTTCAATTCAAATGAAAATAATGATAAAGATTGGTGGGTTGGTCTATGTAATCTTTCAG
AAGTTGGGTTTGGGGTTGATGAGGAGTTGCTTGATTGGGAGTTTCAAGGTAATGTCACTT
GTCAAAGTGATGATCTATGGGATCTCTCAGATATTGGAGAGATAACATTGGAGTGATTGT
ACCGAGCAAGTGGATTGGCGGCCGCTCTAGACAGGCCTCGTACCGGATCTCTAGCTAGAG
CTTTCGTTTCGTATCATCGGTTTCGACAACGTTTCGTCAAGT

>G1640 Amino Acid Sequence (domain in AA coordinates: 14-115)

MGRAPCEKIGLKRGRWTAEEDEILTKYIQTNEGSWRSLPKKAGLLRCGKSCRLRWINY
LRDLKRGNITSDEEEIIVKLHSLGNRWSLIATHLPGRDNEIKNYWNSHLRKYAFT
AVSGDGHNLLVNDVVLKKSCESSSGAKNNNKTKKKKKGRTSRSSMKKHQMVTASQCFSQ
PKELESDFSEGGQNGNFEGESLGPYEWLDGELERLLSSCVWECTSEEAVIGVNDEKVCES
GDNSSCCVNLFEEEQGSETKIGHVGITEVDHDMTVEREREFSFLSSNSNENNDKDWVWGL
CNSSEVGFVDEELLDWEFQGNVTCQSDDLWDLSDIGEITLE*

>G1644 (1..348)

ATGAAATTGATTGGAAAGACTGTGCTTTGATGACTTACACCGAACTCATTTTGGGT
TTCTGCAATGTTTTAATGTTGATCTGCAGGAGGACTAGTGGACCTATGAGACGAGCAAAA
GGTGGTTGGACTCCAGAGGAGGATGAGACACTTAGACGAGCAGTTGAAAAGTATAAGGGG
AAGAGGTGGAAGAAAATAGCGGAATTTTTCCAGAGAGAACACAAGTCCAATGCTTGCAC
AGGTGGCAGAAAGTTCTTAATCCAGAGCTTGTTAAAGGACCTTGGACTCAAGAGGTTCTC
TTATCATTTTCATGTTCTGAACTTTTTTTGGTTTTTCATTTTACGTAA

>G1644 Amino Acid Sequence (conserved domain in AA coordinates: 39-102)

MKLIDWKDCALMTYTELILGFCNVLMLICRRTSGPMRRAKGGWTPPEEDTLRRAVEKYKG
KRWKIAEFFPERTQVQCLHRWQKVLNPELVKGPWTQEVLLSFSCSETFFGFHFT*

>G1646 (34..786)

GATCTTTTGATCCAATCACAAGGCAAGATCCAATGGACAATAACAACAACAACAAC
CAGCAACCACCACCAACCTCCGCTATCCACCTGGCTCCGCCGTCACAACCGTAATCCCT
CCTCCACCATCTGGATCTGCATCAATAGTCACCGGAGGAGGAGCGACATACCACCACCTC
CTCCAGCAACAACAGCAACAGCTTCAAATGTTCTGGACATACCAGAGACAAGAGATCGAA
CAGGTAAACGATTTCAAAAACCATCAGCTCCCTCTAGCTCGTATCAAAAAATCATGAAA
GCTGATGAAGATGTGCGTATGATCTCCGCCGAAGCACCGATTCTCTTCGCGAAAGCTTGT
GAGCTTTTCATTCTCGAACTTACGATTAGATCTTGCTTCACGCTGAAGAGAACAACGCT
CGTACGCTTCAGAAAAACGATATCGCTGCTGCGATTACTAGAACCGATATCTTCGATTTC
CTTGTGATATTGTTCTTAGGGAAGAGATCAAGGAAGAGGAAGATGCAGCATCGGCTCTT
GGTGGAGGAGGTATGGTTGCTCCCGCCGCGAGCGGTGTTCTTATTATTATCCACCGATG
GGACAACCGGCGGTTCTGAGGGGATGATGATTGGAAGACCGGCGATGGATCCTAGCGGT
GTTTATGCTCAGCCTCCTTCTCAGGCATGGCAAAGCGTTTGGCAGAATTCAGCTGGTGGT
GGTGATGATGTGCTTATGGAAGTGAGGAAGTAGCGCCATGGTAATCTCGATAGCCAA
GGGTAAGTGAATTCTAGTAG

>G1646 Amino Acid Sequence (domain in AA coordinates: 72-162)
MDNNNNNNNQPPPTSVPYPPGSAVTTVIPPPPSGSASIVTGGGATYHLLQQQQQQQLQMF
WTYQRQEIEQVNDFKNHQLPLARIKKIMKADEDVRMISAEAPILFAKACELFILELTIRS
WLHAEENKRRTLQKNDIAAAITRTDIFDFLVDIVPREEIKEEEDAASALGGGGMVAPAAS
GVPIYYPPMGQPAVPGGMMIGRPAMDPSGVYAQPPSQAWQSVWQNSAGGGDDVSYSGSGS
SGHGNLDSQG*

>G1672 (239..1399)
CCATTCTGACGTCCGGGATGACGCACAATCCCACTATCCTTCGCAAGACCCTTCCCTCTA
TATAAGGAAGTTTCATTTTCATTTGGAGAGGACACGCTGACAAGCTGACTCTAGCAGATCTG
GTACCGATCACTCCCGTCTTTATCAAATTCCTTCTTCTTCTTACATTTTCCCTATCCAATC
GATCTCACGCAGATCTGATCAATTTCTCATCAAATCATTTAGAGATCAAAAGAAAACCTAT
GAAGAATAGTAAATGTAACCTCATAGATTCAAAGCTCGAAGAATCATCATCTTTGCGG
ATCAAAACATTGTCTGGATGTGGTCGCATGATTCAAGCTGCTACTAAACCAAATTGGGT
TGGATTGCCGGCAGGAGTGAAATTCGATCCGACAGATCAAGAATTATAGAACATTTAGA
AGCAAAAAGTGAAGGGAAAAGAAGAAAATAAGAAATGGTCGTCGTCTCATCCACTTATAGA
TGAATTTATTCCACCATTGATGGAGAAGATGGAATATGTTACACTCATCTCAGAAGCT
TCCAGGGGTGACAAGAGATGGCTTGAGCAAACACTTCTCCACAAACCATCAAGAGCTTA
CACAAACCGGAACAAGAAAACGACGTAAAATAAATCAAACCGATCACGACTCTGAGTTAAC
CGGATCATCAGAAACCGGTGGCACAACCGGGCAAAACAAGACCGGTTATGATCAACGG
TCAACAAAGAGGATGCAAGAAGATATTAGTACTCTACACAAACTTCGGCAAGAATCGTCG
ACCGGAGAAAACAAATTGGGTGATGCATCAATATCATTAGGGATTAATGAGGAAGAGAG
AGAAGGAGAACTTGTGGTCTCCAAGATATTTTATCAGACACAACCAAGACAGTGTGTAG
TAATACTAATTGGTCTGTATCACCATGGTTCCAAGGACGTGATCGGAATTGGTGTGCGAGA
TGAGATTTCCAGCGTAGCTGCCACGTTGCAGAGTCTTGGCTCCGGTGACGTCGTTCTAG
GGTTAATATGCATCCCCATACAAGATCCTTTGATGAGGGGACAGCCGAAGCTTCAAAGGG
AAGAGAGAACCAGCATGTGTCTGGCAGTGCAGGAAGTACATGATGGGATCATAACATC
ATCAATGTCTCATCATATGATTCATGATCATCATAATCAACATCATCAAATCCGAGA
TAGAAGAGAATTTACATGTCTCATCATATCCCATGACCCCTACTATCACATCAACA
TGAGTCAATCTTCCATGTTACAAGTACTATGCCCTTTTCAGCGGCAGCAATTAAGGGGTG
GTCGTCTGGTTCCGGGATTAGAAGACCTAATTATGGGTTGTACCACAGCTACGTGTACAGA
AGACAATAATCACAATGATTAAATTCGCAGGAGCATTGAGAAGCAAACCCCTACGCGAAA
TGCAGAGTGGTTAAGCTTTCCACAATTTCTGGAACCAAGCCGAATCAGATGATCAAAACCG
AAGATTTTAAACAGAACCAAAAGGAAGCAGAGAAATCTTGCAAAAAGCTCCTGCTTAGCTG
TTGATCAATGCCGAAATGCTGAGCTATGACTGACTAGTCTCTGCCATTTAACTTACAAT
ATCACCAGAGGTTGCGATGAATGTTGATTCGCTCAAAGGAGAGCGGCCGCTCTAGACAGG
CCTCGTACCG

>G1672 Amino Acid Sequence (conserved domain in AA coordinates: 41-194)
MKNSKCNLIDSKLEEHHHLGSKHCPGCGRMIQAATKPNWVGLPAGVKFDPTDQELIEHL
EAKVKGKEENKKWSSSHPLIDEFIPTIDGEDGICYTHPQKLPGVTRDGLSKHFFHKPSRA
YTTGTRKRRIQTDHDSLELTGSSETRWHKTGKTRPVMINGQQRGCKKILVLYTNFGKNR
RPEKTNWVMHQYHLGINEEEREGELVVS KIFYQTQPRQCVSN TNWSDHHSKDVIGIGVG
DEISSVAATLQSLGSGDVSVRNMHPHTRSFDEGTAEASKGRENQHVSGTCEEVHDGIIT
SSMSSHMIHDHNQH HQIGDRREFHMS SYPMTPTITSQHESI FHVSTTMPFQRQQLRG
RSSGSGLEDLIMGCTTATCTEDNNHK*

>G1677 (24..1037)
CAGTACTAATTCGTGTGTGTTAATGGTTCTAGTTATGGATGATGAAGAGAGTAACAACG
TTGAAAGATATGACGACGTCGTATTGCCAGGGTTTAGGTTCCATCCCACTGATGAAGAAC
TCGTAAGTTTCTACTTGAAACGGAAGGTTTACACAAATCTCTTCCCTTTGATCTCATCA
AGAAAGTCGACATTTACAAATACGATCCATGGGACCTCCCAAAGCTTGACGCGATGGGGG
AAAAAGAGTGGTACTTTTATTGTCTTAGAGACAGGAAATACCGCAACAGCACAAGACCTA
ACCGAGTAAC TGGAGGTGGCTTCTGGAAGCAACCGGAACAGACCGGCTATATACTCAT
TGGACTCCACTCGATGCATCGGTTTGAAGAAATCACTTGTGTTCTACCGTGGTTCGAGCTG
CTAAAGGAGTCAAAACCGATTGGATGATGCATGAATTCGTCTCCCTTCTCTCTGACT
CTCATCACTCATCATATCCCAATTACAATAACAAGAAGCAACACCTTAACAATAACAACA
ACAGCAAGGAGCTTCCTTCAAACGATGCTTGGGCGATATGTAGAATATTTAAGAAGACAA
ATGCAGTATCCTCACAAAGATCAATCCCAATCTTGGGTTTATCCAACGATTCTCTGACA
ACAATCAACAGTCACACAACAACACCGCAACTCTCTTAGCTTCATCAGACGTTCTCAGCC

ACATATCAACAAGACAAAACCTTTATTCTCTCTCCAGTCAACGAACCCGCAAGCTTCACAG
AATCAGCTGCTTCTTACTTCGCGTCTCAGATGCTCGGAGTCACGTACAATACAGCCAGAA
ACAACGGAACAGGGGATGCTCTGTTTCTGAGAAACAAATGGAACAGGGGATGCTCTGGTTC
TGAGCAACAATGAGAATAACTACTTCAACAACCTTGACTGGAGGGTTGACTCATGAGGTTT
CGAATGTAAGATCAATGGTGATGGAGGAGACTACGGGGAGTGAGATGTCGGCGACGTCGT
ATTCCACTAACAATAAGATCATAGTACTATTAACTTGAATTAGTGTAGACGTTGATC
ATCGCTAATATGTATTAATTTTCTTGTCTTACTATAAACGAAAAA

>G1677 Amino Acid Sequence (conserved domain in AA coordinates: 17-181)

MVLVMDDEESNNVERYDDVLPGRFRFHPTDEELVSFYLRKVLHKSLLPFDLIKKVDIYKY
DPWDLPLAAMGEKEWYFYCPRDRKYRNSRPNRVTTGGGFWKATGTDRIYSLDSTRCIG
LKKSILVYRGRAAGVKTDWMMHEFRLPSLSDSHHSSYPNNYNNKKQHLNNNNNSKELPSN
DAWAICRIFKKTNAVSSQRSIPQSVVYPTIPDNNQQSHNNTATLLASSDVLISHISTRQNF
IPSPVNEPASFTESAASYFASQMLGVTYNTARNNGTGDALFLRNNGTGDALVLSNNENNY
FNNLTGGLTHEVPNVRSMVMEETTGESEMSATSYSTNN*

>G1765 (139..966)

TCCTTCGCAAGACCTTCTCTATATAAGGAAGTTCATTTTATTGGAGAGGACACGCTG
ACAAGCTGACTCTAGCAGATCTGGTACCGTCGACAAGAATGACTTGATTGGTGTTCTAAA
GAGATCGATGTAGTGAAGATGAGTGGCGAAGGTAACCTTAGGTAAGGATCATGAAGAAGAA
AACGAAGCACCCTTCTTGGGTTTCAAGTTTTCATCCGACGGATGAAGAGCTTTTAGGATAC
TATCTTCGAAGAAAAGTAGAGAACAAACCATCAAACCTCGAAGTTATCAAACAGATCGAT
ATCTATAAGTACGATCCTTGGGATCTTCCAAGAGTGAGCAGCGTCGGAGAAAAGGAGTGG
TACTTCTTCTGCATGAGAGGTAGGAAATACAGGAATAGCGTTTCGACCAAACCGAGTGACC
GGTTCAGGTTTCTGGAAAGCCACTGGTATTGATAAACCGGTTTACTCCAATCTTGACTGT
GTTGGTCTCAAGAAATCTCTGGTTTACTATCTTGGTTTCAAGCCGTAAGGCACCAAAACC
GATTGGATGATGCATGAATTCCGCCTCCCTCCACCACGAAAACCGACTCTCCAGCTCAA
CAAGCAGAGGTATGGACACTTTCGAGAATCTTCAAACGAGTCACATCTCAAAGAAACCCA
ACCATCTTACCACCAAACCGAAAACCGGTTATCACTTTAACCGACACTTGTCTAAGACC
AGCAGCTTAGATTCCGACCACAGGACCCGTTACAGTAGATTCCATGTCCCACGAGCCG
CCGCTTCCACAGCCACAGAAATCCTTATTGGAACCAACATATAGTTGGTTTAAATCAACCG
ACATATACTGGTAATGATAATAACCTCCTGATGAGTTTCTGGAACGGCAACGGTGGAGAT
TTCATAGGAGACTCAGCAAGTTGGGATGAACCTAGATCTGTTATAGATGGCAACACTAAA
CCCTAGTAATAAAGTTTCTTTTTCAGCTTTGTACAAAAGATAAAACAAACGGCAACC
GCTCTAGACAGGCCTCGTACCGGATCCTTAGCTAGAGCTTTCGTTTCGTATCATCGGT
TTCGACAACGTTTCGT

>G1765 Amino Acid Sequence (conserved domain in AA coordinates: 20-140)

MSGEGNLGKDHEEENEAPLPGRFRFHPTDEELLYLRRKVENKTIKLELIKQIDIYKYDP
WDLPRVSSVGEKEWYFFCMRGRKYRNSVRPNRVTTGGGFWKATGIDKPVYSNLDVGLKKS
LVYYLGSAGKGTCTDMMHEFRLPSTTKTDSAPQAQAEVWTLCRIFKRVTQQRNPTILPPN
RKPVITLTDTCSTSSLDSDHTSHRTVDSMSHEPPLPQPQNPYWNQHIQVGFNQPTYTGND
NNLLMSFWNGNGGDFIGDSASWDELRSVIDGNTKP*

>G1777 (97..1878)

CTCGTACTTTATCACCTCCGTCGTTCTATAACTCTCTTCCGTCAATCATATCATTTGT
CGACAATTTTATTCTGATCAGTTTAAAAATTGATCCATGGATGATAATTTAAGCGGCGAG
GAAGAAGATTACTATTACTCCTCCGATCAGGAATCTCTCAACGGGATTGATAATGATGAA
TCCGTTTCGATACCTGTTTCTTCCCGATCAAATACTGTCAAGGTTATTACGAAGGAATCA
CTTTTGGCTGCACAGAGGGAGGATTTGCGGAGAGTGATGGAATTGTTATCGGTTAAGGAG
CACCATGCTCGGACTCTTCTTATACATTACCGATGGGATGTGGAGAAGTTGTTTGCTGTT
CTTGTTGAGAAAGGAAAAGATAGCTTGTTTCTGGTGCTGGTGTTACACTTCTTGAAAAC
CAAAGTTGTGATTCTTCCGTTTCTGGTTCTTCTCGATGATGAGTTGTGATATCTGCGTA
GAGGATGTACCGGTTTACAGCTGACAAGGATGGAGTGTGGCCATAGCTTTTGCAATAAC
TGTTGGACTGGGCATTTTACTGTAAAGATAAATGAAGGTCAGAGCAAAAGGATTATATGC
ATGGCTCATAAGTGTAATGCTATTTGTGATGAAGATGTTGTCAGGGCTCTAGTTAGTAAA
AGCCAACCGATTTAGCTGAGAAGTTTGATCGTTTTCTTCTTGAGTCGTATATCGAAGAT
AACAAAATGGTGAAGTGGTGCCGAGTACTCCTCATTGTGGGAATGCCATACGTGTTGAG
GATGACGAGCTCTGTGAGGTTGAATGCTCTTGTGGTTTGCAGTTCTGTTTCAGTTGTTCA
TCTCAAGCTCACTCCCTTGGCTCTTGTGTGATGTGGGAACATATGAGAAAGAAGTGCTTT
GATGAGTCCGAGACTGTTAATTGGATAACTGTTACACAAAGCCGTGTCCCAATGTCCAC

AAGCCTGTTGAAAAGAATGGTGGATGCAATCTCGTGACTTGTCTTTGTGCGACAATCTTTT
TGTTGGTTGTGTGGTGAAGCTACTGGAAGGGACCACACTTGGGCTAGAATCTCGGGTCAT
AGTTGTGGTCGGTTCCAAGAAGATAAAGAGAAACAAATGGAGAGAGCGAAAAGGGATCTC
AAGCGGTATATGCATTATCATAACCGATACAAAGCACATATCGACTCCTCCAAGCTAGAG
GCTAAGCTTAGTAATAATATTAGTAAAAAGGTGTCTATTTAGAAAAGAGGGAGTTACAA
CTTAAAGACTTCAGCTGGGCTACCAATGGACTCCATCGGTATTAGATCAAGACGAGTT
CTTTCATATTATACCTTTTCGCATTTTACATGTTTGGAGATGAGCTGTTTAAAGATGAG
ATGAGCTCTGAGGAAAGAGAAATAAAACAAAATCTGTTTGGAGATCAGCAGCAGCAGCTT
GAGGCTAATGTTGAGAAACTTTCTAAGTCTTGGAGGAACCTTTTGATCAATTTGCTGAT
GATAAGGTCATGCAGATAAGGATTCAAGTCATCAATTTGTCAGTTGCGGTGCGATACACTC
TGCGAAAATATGTATGAATGCATTGAGAATGACTTGTGGGTTCTCTGCAACTTGGCATC
CACAACATTACTCCATACAGATCAAACGGCATAGAACGAGCATCTGATTTTATAGTTCC
CAGAATTCCAAGGAAGCTGTTGGTCAGTCTTCGGATTGTGGATGGACGTCAGGCTCGAT
CAAGCTTTGGAGTCAGGGAAGTCGGAAGACACAAGTTGCTCTTCGGGAAGCGTGCTAGA
ATAGACGAAAGTTACAGAAACAGCCAAACACCTTACTAGATTTAAACTTGCCAGCGGAA
GCCATTGAGCGGAAATGAACACTTATCCTTCTTACCTCCCAATAACACCTTTTGTGTC
AAATAAAGTGTGTTACCCGATATTATAGCTCTAAACCAATCCCCTCTGCTTAATTTG
TCAGTGACCTTACCTAACCTCTTCA
>G1777 Amino Acid Sequence (domain in AA coordinates:124-247)
MDDNLSGEEEDYYYSSDQESLNGIDNDESVSIPVSSRSNTVKVITKESLLAAQREDLRRV
MELLSVKEHHARTLLIHYRWDVEKLFVAVLVEKGKDSLFSGAGVTLLNQSCDSSVSGSSS
MMSCDICVEDVPGYQLTRMDCGHSFCNNCWTHGFTVKINEGQSKRIICMAHKCNAICDED
VVRALVSKSQPDLAKEKFDRLLESYIEDNKMVKWCPSTPHCGNAIRVEDDELCEVECSG
LQFCFSCSSQAHSPSCVMWELWRKKCFDESETVNWITVHTKPCPKCHKPVEKNNGCNLV
TCLCRQSFCLWLCGEATGRDHTWARISGHSCGRFQEDKEKQMERAKRDLKRYMHYHNRYKA
HIDSSKLEAKLSNNISKVSISEKRELQLKDFSWATNGLHRLFRSRRVLSYSYPFAFYMF
GDELFPKDEMSSEEREIKQNLFFEDQQQQLLEANVEKLSKFLEFPDQFADDKVMQIRIQVIN
LSVAVDTLCENMYECIENDLLGSLQLGIHNITPYRSNGIERASDFYSSQNSKEAVGQSSD
CGWTSRLDQALESGKSEDTSCSSGKRARIDESYRNSQTLLDLNLPAAIERK*
>G1793 (59..1783)
AGTGATTTATTGATTAAACCCAAACACAAAATAAACAGATTTGACTCAAAAAGAAGAAAAT
GAATTCTAACAACCTGGCTTGGCTTTCTCTTTACCGAACAACCTTCTTTGCTCCTCA
TGAATACAACCTTGGCTTGGTCAGCGACCATATGGACAACCTTTTCAAACACAAGAGTG
GAATATGATCAATCCACACGGTGGAGGAGGAGATGAAGGAGGAGAGGTTCCAAAAGTGGC
CGATTTTCTCGGTGTGAGCAAAACCGGACGAAAACCAATCCAACCACCTAGTAGCTTACAA
CGACTCAGACTACTACTTCCATACCAATAGCTTGATGCCTAGCGTCCAATCAAACGATGT
CGTTGTAGCAGCTTGTGACTCCAATACTCCTAACAACAGTAGCTATCATGAGCTTCAAGA
GAGTGCTCACAATCTACAGTCACTTACTTTGTCCATGGGGACCAACCGCTGGTAATAATGT
TGTAGACAAAGCTTACAGTCCGAGACCACCGGGGATAACGCTAGCGGTGGAGCACTAGC
CGTTGTTGAGACGGCCACGCCAAGACGTGCATTGGACACTTTCGGACAACGAACCTCGAT
CTATCGTGGTGTCAAGACATCGATGGACTGGTTCGATATGAGGCTCATCTATGGGATAA
TAGTTGTAGAAGGAAGGCCAGTCTAGGAAAGGAAGACAAGTTTACTTGGGTGGATATGA
CAAAGAAGATAAAGCAGCAAGATCATATGATCTAGCTGCACTTAAGTACTGGGGTCCTTC
AACTACTACTAATTTCCCCATTACAAACTACGAGAAAGAAGTAGAGGAAATGAAGCACAT
GACGAGACAAGAGTTTCGTGGCTGCCATTAGAAGGAAAAGTAGTGGATTTTCGAGAGGCGC
TTCGATGTATCGAGGAGTTACAAGGCATCACCAACATGGAAGATGGCAAGCAAGGATCGG
CCGAGTCGCCGGAACAAAGACCTTACTTGGGAACCTTTAGCACTGAGGAAGAAGCAGC
AGAAGCTTACGATATAGCTGCAATAAAGTTTAGAGGACTTAATGCAGTGACCAACTTCGA
GATCAACCGGTACGACGTGAAAGCCATTCTAGAGAGTAGCACTTCCCATCGGAGGAGG
CGCAGCTAAACGGCTCAAAGAAGCTCAAGCTCTTGAGTCTTCAAGGAAACGCGAGGCGGA
GATGATAGCCCTTGGTTCAAGTTTCCAGTACGGTGGTGGCTCGAGCACAGGCTCTGGCTC
CACCTCATCAAGACTTCAGCTTCAACCTTACCCTCTAAGCATTCAACAACCATTAGAGCC
TTTTCTATCTCTTCAACAATGACATCTCTCATTACAACAACAATGCTCACGATTC
CTCCTCTTTAATCACCATAGCTATATCCAGACACAACCTTCATCTCCACCAACAGACCAA
CAATTACTTGACAGCAACGTCGAGCCAGAACTCTCAGCAGCTCTACAATGCGTATCTTCA
TAGCAATCCGGCTCTGCTTCATGGACTTGTCTCTACCTCTATCGTTGACAACAATAATAA
CAATGGAGGCTCTAGTGGGAGCTACAACACTGCAGCATTCTTGGGAACACGGTATTGG

TATTGGGTCCAGCTCGACTGTTGGATCGACCGAGGAGTTTCCAACCGTTAAACAGATTA
CGATATGCCTTCCAGTGATGGAACCGGAGGGTATAGTGGTTGGACAGTGAGTCTGTTCA
GGGGTCAAACCTGGTGGTGTTCCTACTATGTGGAATGAGTAAACAAGGATCTCTTTCTT
CGGCACAAGGAATGGGT

>G1793 Amino Acid Sequence (conserved domain in AA coordinates: 179-255, 281-349)

MNSNNWLGFPPLSPNNSSLPHEYNLGLVSDHMDNPFQTQEWNMINPHGGGGDEGGEVPKV
ADFLGVSKPDENQSNHLVAYNDSYYFHTNSLMPVQSNDVVVAACDSNTPNNSSYHELO
ESAHNLOSLTSLMGTTAGNNVVDKASPSETTGDNASGGALAVVETATPRRALDTFGQRTS
IYRGVTRHRWTGRYEHLWDNSCRREGQSRKGRQVYLGGYDKEDKAARSYDLAALKYWGP
STTTNFPITNYEKEVEEMKHMTRQEFVAAIRRKSSGFSRGASMYRGVTRHHQHGRWQARI
GRVAGNKDLYLGTFTSTEEEAAYDIAAIKFRGLNAVTFEINRYDVKAILESSTLPIGG
GAAKRLKEAQALESSRKREAEMLGSSSFQYGGSSSTSGSTSSRLQLQPYPLSIQPLE
PFLSLQNNDISHYNNNNNAHDSSSPNHHSYIQTQLHLHQQTNNYLQQQSSQNSQLYNAYL
HSNPALLHGLVSTSIVDNNNNNGSSSGSYNTAAFLGNHGIGIGSSSTVSGSTEEFPVTKTD
YDMPSSDGTGGYSGWTSSESQSNPGGVFTMWNE*

>G180 (54..629)

GTAATTACGATCTACAACAAGTGACATCGTCGTCGACGACGATTCAAGAGAATATGAACT
TCCTCGTTCCCTTTTGAAGAAACCAATGTCTTAACCTTTTCTCTTCTTCTTCTCTCTT
CTCTTTCTTCTCTCTTCTTTCCCCATTCACTACTCTTCTCTCCACTACTACTCATGCAC
CTCTAGGGTTTTCTAATAATCTTCAGGGTGGAGGACCTTGGGATCAAAGGTGGTTAATG
ATGATCAGGAGAATTTTGGAGGTGGAACATAAATGATGCTCATTCTAATTCTTGGTGGA
GATCAAATAGTGGAGTGGAGATATGAAGAACAAAGTGAAGATAAGGAGGAACTAAGAG
AGCCAAGATTCTGTTTCCAAACCAAAGCGATGTTGATGTTCTTGACGATGGCTACAAAT
GGCGTAAATATGGTCAGAAAGTCGTCAAGAACAGCCTTCACCCAGGAGTTATTACAGAT
GCACACACAACAACGTAGGGTGAAAAGAGAGTGGAGCGACTATCGGAAGATTGTAGAA
TGGTGATTACTACTTACGAAGGTCTGTCACCAACACATTCCCTCTGATGACTCCACTTCTC
CTGACCATGATTGTCTCTCTTCTTTTAAACATCTCTTCTATATATCTATATATAGACAG
TTATATGTGCACATATAGATGTGTGATATATTGCATATTTGATATTGCATGTGTTTTCA
AGAGTATGTCATCAGATGTTATGCATATATTCTTGACTTGTGCTTATAGTATACATATG
TAATAATATATATTGACATTGGTAGTTTCATTCTGTTCAAACAAAAAAAAAAAAA

>G180 Amino Acid Sequence (domain in AA coordinates: 118-174)

MNFLVPFEETNVLTFSSSSSSSLSSPSFPIHNSSTTTTHAPLGSNNLQGGGPLGSKV
VNDQENFGGGTNNDASNSWRSNSGSDMKNVKIRKLRPRFCFQTKSDVDVLDG
YKWRKYGQKVVKNSLHPRSYYRCHNNCRVKRVERLSEDCRMVITTYEGRHNHIPSDDS
TSPDHDCLSSF*

>G192 (63..959)

CTTTTCTCTTCTCTCCTCAGAGATTGGAAGCTTTTGTCTCCCCTGAGTAACCAATT
CAATGGCCGACGATTGGGATCTCCACGCCGTAGTCAGAGGCTGCTCAGCCGTAAGCTCAT
CAGCTACTACCACCGTATATCCCCCGGCGTTTCTCTCACACAAACCCTATATTACCG
TCGGACGACAAAGTAATGCCGTCTCCTTCGGAGAGATTGAGATCTCTACACACCGTTCA
CACAAGAATCTGTCGTCTCTTCGTTTTCTTGATATAAACTACCCAGAAGAACCTAGAAAGC
CACAGAACCAGAAACGTCCTCTTCTCTCTCTGCTTCTTCCGGTAGCGTCACTAGCAAAC
CCAGTGGCTCCAATACCTCTAGATCTAAAGAAGAAAGATACAGCATAAGAAAGTGTGCC
ATGTAGCAGCAGAAGCTTTAAACTCCGATGTCTGGGCATGGCGAAAGTACGGACAGAAAC
CCATCAAAGGTTACCATATCCAAGAGGATACTACAGATGTAGTACATCAAAGGTTGTT
TAGCCCGTAAACAAGTGGAGCGAAATAGATCCGACCCGAAGATGTTTATCGTCACTTACA
CGGCGGAGCATAATCATCCAGCTCCGACACACCGTAATTCTCTCGCCGGAAGCACACGTC
AGAAACCATCCGATCAACAGACGAGTAAATCTCCGACGACCACTATTGCTACTTATTCAT
CGTCTCCGGTGACTTCAGCCGACGAATTTGTTTTGCCTGTTGAGGATCATCTAGCGGTGG
GAGATCTTGACGGAGAAGAAGATCTGTTATCTTTGTGCGGATACGGTGGTTAGCGATGATT
TCTTCGATGGGTTAGAGGAATTCGAGCCGGAGATAGCTTTCCGGGAACCTCGGCTCCGG
CGAGTTTGTACTCTCTTGGGTTGTGAACAGTGCAGCCGCACTACCAACGGAGGAATATGAT
TAGATTACGACGGCTTAGAATACTCTTATTAGGACAGATTTATAGGATTAAGGAATTATT
CTCGGAGCATATGTAATAATAGGATAAAAGAAAATGTTCTTGTACTTTTTTTTCGGGT
TTCTTCTATTGTTTCTAAACATCTTAGAAAAAATTTAATTGTATATTCCTTAAGCTCGA
TACATCTTGTTTTTAAAAAAAAAAAAAAAAA

>G192 Amino Acid Sequence (domain in AA coordinates: 128-185)

218

219

ALSSSFPSF*

>G265 (280..1317)

CTTTGGTCTTGAAGCCAAATCAAACCTTTCTTCAATCCTCAAATTTTCGAAAATTTTC
TCTTTTGCTTTACGTTCTCTCAATTCCTATTGTAAGAAAGTTTGTTCCTTAATCAATC
AAATCAAAGAGACTTTTGAAGATTGTTTCCCAATTTGCGTCAATCGGGATCGAGTCAAAT
CTGAAATCTTCTCCACTCATCTGACTATAAGACTTAATCAAGGGACTTTTGTTCGG
GTTTGGTTTTAAACGTCTTGGATTGGAAGTGGTTAAGGTATGGATGAAAATAATGGAGGT
TCAAGCTCACTTCCACCTTTCTTACTAAAACATATGAAATGGTTGATGATTCTTCTCT
GACTCGGTCTGTTGCTTGGAGCGAAAACAACAAAGCTTCATCGTCAAGAATCCAGCAGAG
TTTTCAAGAGACCTTCTTCCGAGATTCTTCAAGCATAAGAATTTCTCAAGTTTCATCCGT
CAGCTTAATACATATGGTTTTTCGAAAAGTAGATCCTGAGAAATGGGAATTCCTGAATGAT
GATTTTGTTAGAGGTGACCTTACCTTATGAAGAACAATTCATAGACGAAAACCGGTTTCAT
AGCCACTCGTTAGTGAATCTACAAGCGCAAAATCCCTTGACGGAATCAGAAAGACGGAGC
ATGGAGGATCAGATAGAAAGACTGAAAAATGAGAAAGAAGGCCTTCTTGGCGAGTTACAG
AACCAAGAGCAAGAACGGAAAAGAGTTTGAGCTGCAAGTAACGACATTGAAAGATCGGTTA
CAACATATGGAACAACATCAGAAATCAATAGTGGCATATGTTTCACAGGTTTGGGAAAA
CCAGGACTTTCACTAAACCTCGAAAACCATGAGAGAAGAAAAAGAAGATTCAAGAGAAC
TCTCTTCTCCAAGCAGTTACACATAGAACAGGTGCAAAAGTTAGAATCTTCGCTAACG
TTTTGGGAGAATCTTGTATCGGAATCATGCGAGAAGAGCGGTTTGCAGTCATCAAGCATG
GATCATGATGCAGCTGAGTCAAGTCTAAGTATTGGCGATACACGACCCAAATCATCGAAG
ATTGATATGAACCTCAGAGCCGCCCGTTACCGTTACTGCGCCTGCTCCAAAAACAGGCGTT
AACGATGACTTTTGGGAACAATGTTTGACAGAGAACCCTGGATCAACCGAGCAACAAGAA
GTTCAAGTCAGAGAGAAGAGATGTCGGTAATGATAATAATGGTAATAAGATTGGAAATCAA
AGGACGTATTGGTGGAAATTCAGGGAATGTAAATAACATTACAGAGAAAGCTTCTTGACAT
GAATGAGGTTTTTGTAAATAAGTTTTCTTTTGGTTCCACTGAGATTATTGTATGTGTTCA
TTATTTATTACTCTGTTTCTGTAAAAACAAATCTCTCTATTGTTTGAGGCAGGAGTGACA
TAAATGCATATGCAGAATTGGTTTCAAAA

>G265 Amino Acid Sequence (domain in AA coordinates: 11-105)

MDENNGSSSLPPFLTKTYEMVDDSSSDSVVAWSENKSFIVKNPAEFSRDLLPRFFKHK
NFSSFIRQLNTYGRKVDPEKWEFLNDDFVRGRPYLMKNIHRRKPVHSHSLVNLQAQNPL
TESERRSMEDQIERLKNKEGLLAELOQEQERKEFELQVTTLDRLQHMEQHQKSIVAY
VSQVLGKPLSLNLNENHERRKRRFQENSLPPSSSHIEQVEKLESSLTFWENLVSECEKS
GLQSSMDHDAESSLSIGDTRPKSSKIDMNSEPPVTVTAPAPKTGVNDDFWEQCLTENP
GSTEQQEVQSERRDVGNDNNGNKIGNQRTYWWNSGNVNNITEKAS*

>G2792 (1..960)

ATGGATCATCATCATCATAGCATCAAGAAATTCATCAACAACATCAGAATTACCATCA
TTCGAGCCAGCGTGCCATAACGGTAATGGTAACGGTTGGATCTATGACCCAAATCAAGTT
AGGTACGATCAAAGTAGTGACCAACGGCTGTCAAAGTTGACGGATCTTGATAGGCAAGCAC
TGGTCAATTGCACCACCGAATAATCCCGACATGAACCATAACCTTCATCATCACTTCGAT
CATGATCATTCTCAAACGACGACATTTCTATGTACAGACAAGCCTTGGAGGTGAAAAAT
GAGGAAGATCTTTGTTACAATAATAGGCTCAAGTGGTGGTGGTTCTTGTTCATGATCCT
ATAGAAAGTTCTAGAAGTTTCTTGTATATAAGGTTAAGTAGGCCATTAACGATATTAAAT
CCGTCAATTAAGCCATGCTTTAAGGCCTTAAACGTATCCGAGTTTAAACAAGAAAGACAT
CAAACGGCATCTCTGGCAGCAGTGAGACTGGGAACAACAACCGTGGAAAAAGAAGAGA
TGTGAAGAAATTTCCGATGAGGTTTCAAAGAAGGCCAAGTGCAGTGAGGGCTCTACACTT
TCGCCAGAGAAGGAACTACCCAAAGCCAACTTCGAGACAAGATCACGACTCTACAGCAA
ATTGTGTCTCCCTTTGGAAAGACTGATACTGCTTCTGTGCTTCAAGAGGCCATCACTTAC
ATAAATTTTTATCAAGAGCAAGTTAAGCTGCTAAGCACTCCTTATATGAAGAATTCATCA
ATGAAGGATCCATGGGGGGGATGGGACAGAGAAGATCACAACAAAGGGGACCGAAGCAT
CTTGATCTAAGGAGTAGAGGGCTTTGTTTGGTTCTTATTTCATATACCCCAATCGCATAC
CGCGATAACAGTGCAACTGACTACTGGAATCCACGTATAGAGGTTCTTTGTATCGTTAG

>G2792 Amino Acid Sequence (domain in AA coordinates: 190-258)

MDHHHHIASRNSSTSELPSFEPACHNGNGNGWIYDPNQVRDQSSDQRLSKLTDLVGKH
WSIAPPNNPDMNHNLHHHFDHDSQNDISMYRQALEVKNEEDLCYNNGSSGGGSLFHDP
IESSRSFLDIRLSRPLTDINPSFKPCFKALNVSEFNKKEHQ TASLAAVRLGTTNAGKKR
CEEISDEVSKKAKCSEGSTLSPEKELPKAKLRDKITTLQQIVSPFGKTD TASVLQEAIT
INFYQEQVKLLSTPYMKNSSMKDPWGGWDREDHNRGPKHLDLRSRGLCLVPISYTPAIY

RDNSATDYWNPTYRGSLYR*

>G2830 (1..903)

ATGTCTTCCATCCCAAATAGGTTCAATATTTATGGTGGTGATACCACAAACCATCGTGAA
TCGCTTCCCATCGAAATGAATCACAACCTCTCGAATGGTTCGATCCATGTTTCATTACATCT
GATCGCATGAATCATAGAGATTTGTTTTCTTCTCCTCCTTCTTCTTCTTATCAAAAT
TCACATATCTTCTCATCTTCTGTTGGGTTTAATAATTACATATGACTTATCATATGCTG
AAAAGAAATTATGATTCTGTTTCCCGTGTGATTATTTCTCTACTAAAGATCATTCTCAT
TTTACTCAAGTATCTTTCACCTCAAACCATCACAAATAAGTATACTACTATTGTTCCCTTCC
AATATATTTGACACTGTTCACTATGATATTGGTTCGTGTCAAACGTGCCATAGATTTTAGA
AATATTTGGAATCTTAAATCTCATCTTCCAAAAAATTTAATAGGCAATGCGAGATTTTG
AATCCTACCCCTCTTAATATCGTCTTTCGCGACCAGGATTCAGCTGATCGTCAACATTTA
GACATTATTTTCTCGTCATCAAAGCACAAACCATGTTTTCGAAGATGGTTCGATCCTTGAAG
AAAATTTCCGAACCAACCAATCTGTTTGAAAAATCTAATTCTTATGATTCTCAAGAAGAT
GAGAAATCGATGCTTATCAATATGATGGTTCGTACACATAGTCTACCGTATACGAAATAC
GGTCCATATACATGTCCAGGTGTAACGGTGTGTTTGATACTTCTCAAAAAATTTGCTGCA
CATATGTTATCTCACTACAATAATGAGACGGACAAAGAAAGAGACCAAAGATTTTCGTGCA
AGAAATAAAAAACGATATCGTAAGTTTATGGACAGTCTTAAATATCAAAACAGAAGATA
TGA

>G2830 Amino Acid Sequence (domain in AA coordinates:245-266)

MSSIPNRFNIYGGDTTNHRESLPIEMNHNMRMVRSMFITSDRMNRDLFSSPPSFSSYQN
SHISSSSVGFNNSHMTYHMLKRYDSVSRADYFSTKDSHFTQVSFTQTI TNKYTTIVPS
NIFDTVHYDIGRVKRAIDFRNIWNPKSHLPKKFNRQCEILNPTPLNIVFPHQDSADRQHL
DIIFFSSSKHNHVFDGRSLKKISEPTNLF EKSNYSYDSQEDEKIDAYQYDGRTHSLPYTKY
GPYTCPRCNGVFDTSQKFAAHMLSHYNNETDKERDQRFARNKKRYRKFMDSLKISKQKI

*

>G286 (94..2454)

TGCAATTTCTCTCGACCAAACCTAATTTTCAGGTTTGGGGTTTTCCTTCTTTCCTGTC
AATTTTGATGAACTTGTGATTAGTATGAAATGAATGCTAATGAGCAAACCTCGATCC
GCCAATGGCATTGGCAATGGCAATGGTGAAGTCTATTCCCGGGATTCCAGATGACTTACGG
TGCAAGAGATCGGATGGTAAACAGTGGAGATGCACTGCAATGTCCATGGCTGATAAGACT
GTTTGTGAGAAGCACTACATCCAAGCAAAGAAGCGGGCGGCTAATTTCTGCTTTCAGGGCG
AACCAGAAGAAAGCGAAAAGCGGATCATCGTTAGGCGAAACAGATACGTATTTCGGAAGGG
AAGATGGATGATTTTCGAGTTACCAAGTACCAGTACCAGTACCAGTACCAGTACCAGTACCAGT
TCTGCTTCCAAGAGTAATGGTAGACTAGAGAAAGAGACATAATAAAAGCCTGATGCGGTAC
TCGCGCGAGACACCGATGATGAGGAGTTTCTCTCCACGTGTTGCAGTGGATTTGAATGAT
GACTTGGGTAGAGATGTTGTAATGTTTGAAGAGGGCTACAGATCTTATAGGACACCACCA
TCTGTTGCTGTTATGGATCCGACACGAAACAGATCACACCAAAGCACCAGTCCATGGAA
TACTCAGCAGCAAGCACAGATGTGTCTGCAGAGTCTTTGGGGGAAATCTGCCATCAATGC
CAGAGAAAAGATAGAGAGAGAATCATTTCTTGCCCTCAAATGCAATCAAAGAGCCTTCTGC
CACAAATGTCTATCGGCAAGGTACTCGGAGATATCACTTGAAGAAGTCGAGAAAGTTTGC
CCTGCATGTCGTGGCTTGTGTGATTGCAAATCTTGCCCTGCGTTTCAGATAATAACAATAAG
GTTCCGATCCGGGAAATACCCGTTTTTGGACAAGTTGCAGTATCTTTATCGTCTATTATCA
GCTGTCTTACCAGTCATAAAGCAGATCCATCTTGAACAATGTATGGAAGTTGAACTAGAG
AAGAGGCTTCTTGAAGTTGAGATTGATCTTGTTCAGGGCAAGATTGAAAGCAGATGAGCAG
ATGTGCTGCAACGTGTGTTCGGATACCAAGTGTGACTACTACCGTCACTGTCCGAAGTGC
TCATATGACCTTTGCCTGAGATGCTGTCAAGATCTACGGGAAGAGTCTTCAGTGACGATT
AGTGGGACTAACCAAAACGTACAAGATAGAAAAGGAGCTCCCAAACCTAAACTAACTTT
TCATACAAGTTTCTTGTGAGTGGGAAGCCAACGGTGTATGGGAGCATCCCTTGCCCTCCTAAG
GAGTATGGAGGCTGCGGTTACATCTTTGAATCTTGCCCGCATTTTCAAGATGAATTGG
GTTGCAAAGCTTGTGAAAAATGCTGAGGAGATTGTTAGTGGCTGCAAATTATCTGATCTT
CTGAACCTGATATGTGTGATTCAAGATTCTGCAAATTTGCTGAGAGAGAAGAGAGCGGT
GACAACTACGTGTACAGCCCGTTCGCTTGAAACGATTAATACTGATGGAGTAGCTAAGTTT
GAGCAACAATGGGCGAGAGGGTTCGCTTGTACTGTGAAATGGTACTTGATGACTCATCT
TGCTCTAGATGGGATCCTGAGACTATTGAGAGGATATAGACGAGCTTTTCGGACGAGAAA
CTGAGAGAACATGATCCATTCTTGAAGGCCATTAATTGCTTGGATGGTTTAGAGGTTGAT
GTAAGACTTGGGGAGTTTACAAGAGCATATAAAGATGGAAGAACCAAGAGACAGGTCTT
CCGCTATTGTGGAAGTTAAAGGACTGGCCGAGCCCAAGTGTCTCCGAGGAGTTTCAATTTT

TACCAAAGACCTGAGTTTATCAGAAGTTTCCGTTTCTCGAGTACATTCATCCCCGGTTA
GGCCTTCTGAATGTTGCAGCCAAGTTACCTCATTACTCGCTCCAAAACGATTTCAGGTCCA
AAGATTTATGTGTCTTGTGGGACGTACCAAGAAATCAGTGCTGGCGATTTCATTGACTGGT
ATTCACTACAACATGCGTGACATGGTATACCTATTGGTGACACAGTCTGAAGAAACAACA
TTCGAAAGGGTGAGAAAAACAAACCTGTTCCAGAGGAACCTGACCAGAAGATGAGCGAA
AATGAGTCACTTCTTAGCCCTGAGCAGAAATTAAGGGACGGAGAGTTACATGATCTATCA
CTTGGTGAAGCCAGTATGGAGAAGAATGAACCTGAGTTGGCGTTGACTGTGAATCCAGAG
AACTTAACGGAACCGGTGACAACATGGAATCTTCTTGACATCTTCATGTGCAGGAGGA
GCCCAGTGGGATGCTTTTCGACGCCAAGACGTCCCAAAGTTGTCCGGGTATTTGCAGAGA
ACATTCCAGAAGCCTGATAATATCCAGACTGATTTTGTAAAGCCGTACCTGCTAATTCAA
TAAATGAAGTGTGTAAAGTCTTGTATGTGGAATGATTGAGTTTCCTAGTTTGTCTTACTCT
GGTTTCAGGTGTACGCCCCGTGTGTATGAAGGATTGTCTTTAAATGAACACCACAAGAGAC
AACTAAGAGACGAGTTTGGAGTTGAGCCATGGACATTTGAGCAACATCGTGGTGAGGCTA
TCTTCATTCCGGCTGGATGTCCGTTCCAAATCACTAATCTTCAGTCGAATATTCAGGTGG
CACTTGACTTCTTGTGCCCTGAAAGCGTTGGAGAGTCAGCAAGACTAGCTGAAGAAATCC
GGTGTTTACCAAACGACCACGAGGCAAACTTCAGATTCTAGAGATTGGAAGATATCAT
TATACGCAGCTAGCTCAGCCATTAAAGAGGTTTCAGAACTGGTCTTGGATCCAAAGTTTG
GAGCAGAGCTTGGATTTGAAGACTCTAACTTAACCAAAGCAGTCTCTCACAACCTTAGACG
AGGCAACCAAGCGGCC

>G286 Amino Acid Sequence (domain in AA coordinates: TBD)
MNANEQTRSANGIGNNGESIPGIPDDLRCRSDGKQWRCTAMSMADKTVCCKHYIQAKK
RAANSAFRANQKAKRRSSLGETDTYSEGKMDDFELPVTSIDHYNNGLASASKSNGRLEK
RHNKSLMRYSPEPMMRSFSPRVAVDLNDLGRDVMFEEGYRSYRPPSVAVMDPTRNR
SHQSTSPMEYSAASTDVSABSLGEICHQCQRKDRERIISCLKCNQRAFCHNCLSARYSEI
SLEEVEKVCAPACRGLCDCKSLRSDNTIKVRIREIPVLDKLYLYRLLSAVLPVIKQIHL
EQCMEVELEKRLLEVEIDLVRARLKADEQMCCNVCRIPVVDYRHCPCNSYDLCLRCCQD
LREESSVTISGTNQNVDKRGAPKLKLNFSYKPEWEANGDGSIPCPKEYGGCGSHSLN
LARIFKMNWVAKLVKNAEEIVSGCKLSDLLNPDMDCSRFCFAEREESGDNYVYSPSLET
IKTDGVAKFEQQWAEGRIVTVKMLVDDSSCSRWDPETIWRDIDELSDKLREHDPFLKAI
NCLDGLLEVDRLEGEFTRAYKDGKNQETGLPLLWKLKDWSPSASEEFIFYQRPEFIRSF
FLEYIHPRLGLLNVAAKLPHYSLQNDSGPKIYVSCGTYQEISAGDSLTIHYNMRDMVYL
LVHTSEETTFERVRKTKPVPEEPDQKMSSENLSPQKLRDGLHDLGLGEASMEKNEP
ELALTVNPNELTENGDNMESSCTSSCAGGAQWDFRRQDVPKLSGYLQRTFOKPDNIQTD
FVSRTC*

>G291 (124..1197)
CAAGAACCCAAAGATCTCTCTCTATTTGTTTGCCTTCTTCTTTCTTTCTGACTCAAACCC
TCAAATCAATTCTCGCGATTAAGCAAAACCCTAGATTTATTCTACTCTTCGAAGTCGATT
TCAATGGAAGGTTCTCTCGTCAGCCATCGCGAGGAAGACATGGGAGCTAGAGAACAACATT
CTCCAGTGGAACCAACCGATTACGCTCCGACAGTATATTCCACTACGACGACGCTTCA
CAAGCCAAAATCCAGCAGGAGAAGCCATGGGCCTCCGATCCTAACTACTTCAAGCGCGTT
CACATCTCAGCCCTTGCTCTTCTCAAGATGGTGGTTACGCTCGCTCCGGTGGCACAATC
GAGATCATGGGCTTTATGCAGGGTAAACCGAGGGTGATACAATCATCGTTATGGATGCT
TTTGCTTTGCCTGTTGAAGGTACTGAGACTAGGGTTAATGCTCAGTCTGATGCCTATGAG
TATATGGTTGAATACTCTCAGACCAGCAAGCTGGCTGGGAGGTTGGAGAACGTTGTTGGA
TGGTATCACTCTCACCTTGGGTATGGATGTTGGCTCTCGGGTATTGATGTTTCGACACAG
ATGCTTAACCAACAGTATCAGGAGCCATTCTTAGCTGTTGTTATTGATCCAACAAGGACT
GTTTCGGCTGGTAAGGTTGAGATTGGGGCATTTCAGAACATATCCAGAGGGACATAAGATC
TCGGATGATCATGTTCTGAGTATCAGACTATCCCTCTTAACAAGATTGAGGACTTTGGT
GTACATTGCAACAGTACTACTCATTGGACATCACTTATTTCAAGTCATCTCTCGATAGT
CACCTTCTGGATCTCTTTGGAACAAGTACTGGGTGAACACTCTTCTTCTTCCCCACTG
TTGGGCAATGGAGACTATGTTGCCGGGCAAATATCAGACTTGGCTGAGAAGCTCGAGCAA
GCGGAGAGTCAGCTCGCTAACTCCCGGTATGGAGGAATTGCGCCAGCCGGTCACCAAAGG
AGGAAAGAGGATGAGCCTCAACTCGCGAAGATAACTCGGGATAGTGCAAAGATAACTGTC
GAGCAGGTCCATGGACTAATGTACAGGTTATCAAAGACATCTTGTTCATTCCGCTCGT
CAGTCCAAGAAGTCTGCTGACGACTCATCAGATCCAGAGCCCATGATTACATCGTGAAGT
TGGTCTATTCTTTTGGTCTGCGGAAATTGACTATCGGTTTGACCCGGTTTATGA
GGCAATGCCATTGTTCCCTATATCTCTAGTGTAGTATCTGCTTCAGACAAAGATCTTTG

GGTTATTAAATGACATTAACATAAAAAAA
>G291 Amino Acid Sequence (domain in AA coordinates: 132-160)
MEGSSSAIARKTWELENNILPVEPTDSASDSIFHYDDASQAKIQQEKPWASDPNYFKRVH
ISALALLKMOVHARSGGTIEIMGLMQGKTEGDTIIVMDAFALPVEGTETRVNAQSDAYEY
MVEYSQTSKLAGRLNVVGVYHSHPGYGCWLSGIDVSTQMLNQQYQEPFLAVVIDPRTV
SAGKVEIGAFRTYPEGHKISDDHVSEYQTIPLNKIEDFGVHCKQYSLDITYFKSSLDLSDH
LLDLLWNKYVWNTLSSSPLLNGNDYVAGQISDLAEKLEQAESQLANSRYGGIAPAGHQRR
KEDEPQLAKITRDSAKITVEQVHGLMSQVIKIDILFNSARQSKSADDSDDPEPMITS*
>G427 (49..1230)
TTTCCCTCTCCGAAACAGAAATTCAAAAACAAATTCACACGAAAACGATGGCGTTTCAT
AACAAATCACTTTAATCATTTCACCGACCAACAACAACATCAGCCTCCTCCTCCGCCGCAA
CAGCAGCAGCAACAACATTTTCAAGAATCAGCACCCCTAATTGGCTCCTCCGCTCCGAC
AACAACTTCTCAATCTCCACACAGCTGCCACAGCCGCCGCTACAAGCTCCGATTCTCCT
TCTTCCGCCGCCGCTAACCACTGGCTCTCACGATCCTCATCCTTCTTCCAACGAGGCAAC
ACCGCAAACAACAACAACAACGAAACATCCGGTGACGTCATCGAAGACGTTCCCGCGCGA
GAGGAGTCAATGATCGGAGAGAAGAAGGAGGCGGAGAGGTGGCAGAATGCCGAGACACAAG
GCGGAGATACTGTCTCATCCACTATACGAGCAACTTTTGTGCGGCACACGTGGCGTGCCTG
AGGATCGCAACGCCGGTGGATCAGCTTCCGAGGATAGACGCACAGCTTGCTCAGTCTCAA
AACGTCGTGGCTAAGTACTCAACTTTAGAAGCCGCTCAAGGACTCCTCGCCGGCGATGAC
AAGGAGCTTGACCATTTCATGACGCATTATGTACTATTGCTTTGCTCTTCAAAGAACAA
CTGCAACAGCATGTTTCGTGTTTCATGCAATGGAAGCTGTTATGGCCTGTTGGGAGATTGAA
CAGTCGCTTCAAAGTTTACAGGAGTATCTCCTGGTGAAGGCACAGGAGCAACAATGTCT
GAGGATGAAGATGAGCAAGTAGAGAGTGATGCTCATTTGTTTGATGGAAGCTTAGATGGG
TTAGGGTTTGGTCTCTAGTTCCCACTGAGAGCGAGAGATCTTTGATGGAACGAGTCAGA
CAAGAACTCAAACATGAACCTCAAGCAGGTTTACAAGGAGAAAATTGTGGACATAAGAGAG
GAGATACTGAGGAAGAGAAGAGCTGGAAAATTACCAGGAGACACCACCTCTGTTCTCAA
TCATGGTGGCAATCTCATTTCTAAGTGGCCTTACCCTACTGAGGAAGATAAGGCGAGGTTG
GTGCAGGAGACGGGTTTGCAGCTCAAACAGATAAACAATTGGTTCATCAATCAAAGAAAG
AGGAATTGGCATAGCAATCCATCTTCTTCTACCGTCTCAAAGAATAAACGCCGAAGCAAT
GCAGGTGAAAACAGCGGAAGAGACCGTTGAGATCAAGCTTGATGTAGAGATCCAAAAGC
TTTATAGAAAGGTGGAGGCATGAAGACAAAGAATTCTTACACAACAACGTAGGACGTAA
TTTTGTGCCAGTACATGGTATGGCTTTTCATATTTGGTAATGATTAGGGCCACACAAAATT
AAACCCCAAAGCATGATTTGTAATATGAGGTTTTAGATGGACTTTATGATAGGATCGTCA
GTCTTCACTGCCATCTCCATTCTCCACCATCAATCCATCATTATATCTTTGTGAAAAAAA
A
>G427 Amino Acid Sequence (domain in AA coordinates: 307-370)
MAFHNNHFNHFTDQQQHQP PPPPQQQQQHQH FQESAPPNWLRLSDNNFLNLHTAATAAATS
SDSPSSAAANQWLSRSSSFLQRGNTANNNNNNETSGDVI EDVPGGEESMIGEKKEAERWQN
ARHKAEILSHPLYEQLLSAHVACLRIATPVDQLPRIDAQLAQSQNVVAKYSTLEAAQGLL
AGDDKELDHFMTHYVLLLCFKEQLQQHVRVHAMEAVMACWEIEQSLQSF TGVSPGEGTG
ATMSEDEDEQVESDAHLFDGSLDGLGFGPLVPTESERSLMERVVRQELKHELKQGYKEKIV
DIREEILRKRRAGKLPD TT SVLKSWWQSHSKWPYPTEEDKARLVQETGLQLKQINNWF I
NQRKRNWHSNPSSSTVSKNRRSNAGENSGRDR*
>G509 (122..1054)
CTTCTCTCTTTGCTAATAAACTTTTCTTTGAACCTTACACGCCTTGTTGATATTACTCTC
TTAAATATATATTTTCGTACATTACACAGACATATATAAAGCTAAAGATTCTTCACGT
AATGGGTTTGAAAGATATTGGGTCCAAATTGCCACCGGGGTTTCGATTTTCATCCAAGTGA
TGAAGAGTTGGTTTTCATTATCTTTGCAACAAGATTAGGGCCAAATCTGATCATGGTGA
TGTTGATGATGATGATGATGATGTTGATGAAGCTTTGAAGGGTCTACTGATCTTGTGGA
GATTGACTTGATATCTGTGAGCCATGGGAGCTTCTGATGTGGCAAAGTTAAACGCAAA
GGAATGGTACTTCTTCCAGTTTCCGTGATCGAAAGTATGCTACTGGATATCGCACGAACAG
AGCGACAGTAAGCGGATACTGGAAAGCAACAGGAAAAGATCGAACGGTGATGGATCCACG
TACAAGGCAATTGGTAGGGATGAGAAAAACACTAGTGTCTACAGAAACAGAGCACCAAA
TGGGATCAAACTACTTGGATCATGCACGAGTTCCGTCTTGAGTGTCTTAACATCCACA
TAAGGAAGACTGGGTCTTGTGCAGAGTGTTCACAAAGGCAGAGACTCATCGCTACAAGA
CAATAATTATTATAACAATGATAATCAGACGCAAGGCTTGAAGTTAATGACGCTCCGGA
TCTTAATTACAACAATCAGTTGCCACCTTTGCTATCATCCCTCCTCATAATCATCAACA

TGAGAAGATGAAAATCCAAGTTTGTGATCAGTGGGAGCAGCTAATGAAGCAGCCTTCAAG
GACCACCGGCCACCCCTATCATCACCATTGTTCATCATCAAACCATAGCATGTGGTTGGGA
GCAGATGATGATCGGTTTCGCTGTTCATCACCTTCGAGTCATGGCCCTGATCACGAGTCCTT
TGCTAAATTTGCTTTACCGTCGACAATAACAACAGTGTCAACATCAGTGGTGATCATCAT
CAGAATTATGAGAAGATTTTGTGTTCATCACTAGACATGACGAGTTTGGATCACGACAAG
ACATGTATGGGATCATCATCGGATGGTGGTATGGTCTCTGATCTTCACATGGAATGTGGT
GGATTGAGTTTGGAGACCGAAAATATCCTCGCTTTCCAATGAACATAATTCAAGGGGTTT
GCCAATTTGTTGATTCGTGAATTATACAAACATTTTATCTATAGATTTATCACATTATCA
AACATGTAAAGTTGTGTGGCATTGGGTATAGGGTTTGTGTGATTCTAGGTTTATAGGACG
ATGTATGTTGTTATATTTAGCGTGTTTTAGGATTTATTCTCATTTTAAATATATGAA
AACCATTACTATGAATACAATTAGTTTCTTTGTGTAAATAATATTTTAGATTATCAA
AAAAAAAAAAAAA

>G509 Amino Acid Sequence (domain in AA coordinates: 13-169)
MGLKDIGSKLPPGFRFHPSEELVCHYLCNKIRAKSDHGDVDDDDDDVDEALKGSTDLVE
IDLHICEPWELPDVAKLNAKEWYFFSFRDRKYATGYRTNRATVSGYWKATGKDRTVMDPR
TRQLVGMRLTLVFYRNRAPNGIKTTWIMHEFRLECPNIPHKEDWVLCRVFNKGRDSSLQD
NNYYNNDNQTRLEVNDAPDLNYNQLPPLLSSPPHNHQHEKMKIQVCDQWEQLMKQPSR
TTGHPYHHCHHTIACGWEQMMIGSLSSPSSHGPDHESFAKFALPSTITTVSTSVVII
RIMRRFCCHH*

>G519 (85..894)
CACAAAGATCCTCCGATTGGAAGGTTTATAAAACTCAAATCGAATCTTATCCACAAGA
AAACAACAAGGTACTTTTCCAAAATGAAGCGGAGTTGAATTTGCCGGCGGGATTCCGA
TTTCATCCGACGACGAAGAGCTTGTCAAGTCTATCTTTGCCGGAGATGTGCGTCAGAA
CCGATTAAACGTTCCGGTTATCGCAGAGATTGACTTGTACAAATTCAATCCATGGGAGCTT
CCAGAAATGGCGTTGTACGGTGAGAAAGAATGGTACTTCTCTCGCATAGAGACCGGAAA
TACCCAAACGGGTCGAGACCAAACCGGGCAGCTGGAACCGGTTATTGGAAGCGACTGGA
CTGTATAAACCGATCGGAAAACCGAAGACGTTAGGGATTAAGAAAGCACTCGTCTTCTAC
GCAGGAAAAGCTCCGAAAAGGGATTAAACGAATTGGATTATGCACGAGTATCGTCTCGCT
AATGTGATCGATCGATCTGCTTCTACCAACAAGAACAAGCAACTTAAGACTTGATGATTGGGTT
TTGTGTGCGATATACAATAAGAAAGGAACAATGGAGAAGTATTTACCGCGCGCGCTGAG
AAACCGACGAAAAGATGAGTACGTCCGACTCAAGATGCTCAAGTCACGTGATTTACCG
GACGTACAGTGTCTGTAACTGGGAGGTTGAGAGTGAGCCCAATGGATTAATCTGGAA
GACGCGTTAGAGGCATTTAATGATGACACGTCCATGTTTAGTTCCATTGGTTTGTGCAA
AATGACGCCTTTGTTCTCAGTTTTCAGTACCAGTCCCTCCGATTTCTGTCGATTCTGTTT
GACCCGTTTCGAGCAGAAACCGTTCTTGAATTGGAATTTTGCTCCTCAAGGGTAAAAATAA
TCGGCAAAAAGTTGAAGCTTTTCAGAGTCTTCGATCACCGGCATTGTGTGCGATCCTGAC
CCGGAGACCAAGTCGGCTCATACGATTACATAATCGGGTTATTGAGATTTCCACATTTGG
ATTTCCGAGACTAACCAACTTAACGGATTCTGGGGTAATTGGGGGGTTTTCACAGGTGA
ATCAGACTGAGTCAGCAAGTTTTCGATTTTGGTTTTGTGTTTGAATGATTGATTAAATG
TCTAAAGATATCAGCAAGTAGATTGAGAAGAACTGTAAGCAATTGTGACCACCGTTA
TGAATCATAAATATATCAATGAAGCATGAGCTTATTTTTTTTTTAAAAA

>G519 Amino Acid Sequence (conserved domain in AA coordinates: 11-104)
MKAEELNLPAGFRFHTDEELVKFYLCRRCASEPINVPVIAEIDLYKFNPWELPEMALYGE
KEWYFFSHRDRKYPNGSRPNRAAGTGYWKATGADKPIGPKTLGIKKALVYAGKAPKGI
KTNWIMHEYRLANVDRSASTNKKNNLRLLDDWVLCRIYNKKGTMKYLPAAAEKPTKEMST
SDSRCSSHVISPVDTCSDNWEVESEPKWINLEDALEAFNDDTSMFSSIGLLQNDAFVPOF
QYQSSDFVDSFQDPFEQKPFNLWNFAPQG*

>G561 (86..1168)
AATTTGTTTTTTTTTCTTTTGTGGGTTCAATTGGAATTGTTTCCCTGAGACTCAAGTTA
CTGTGTCACTACTCTGCATTGAGCAATGGGTAGCAACGAAGAAGGAAACCCACTAACAA
CTCTGATAAGCCATCGCAAGCTGCTGCTCCTGAGCAGAGTAATGTTTATGTGTATCATCA
TGACTGGGCTGCTATGCAGGCATATTATGGGCCTAGAGTTGGTATACCTCAATATTACAA
CTCAAATTTGGCGCCTGCTCATGCTCCACCGCCTTATATGTGGGCGTCTCCATCGCCAAT
GATGGCTCCTTATGGAGCAGCATATCCACCATTTTGCCCTCCTGGTGGAGTTTATGCTCA
TCCTGGTGTTCAAATGGGCTCACAAACCACAAGGTCCTGTTTCTCAATCAGCATCTGGAGT
TACAACCCCTTTGACCATTGATGCACCAGCTAATTCAGCTGGAAACTCAGATCATGGGTT
CATGAAAAGCTGAAAGAGTTTCGATGGACTTGCAATGTCAATAAGCAATAACAAAGTTGG

GAGTGCTGAACATAGCAGCAGTGAACATAGGAGTTCTCAGAGCTCCGAGAATGATGGCTC
TAGCAATGGTAGTGTGGTAATACAACCTGGGGGAGAACAACTAGGAGGAAAAGAAGGCA
ACAAAGATCACCAAGCACTGGTGAAAGACCCCTCATCTCAAAACAGTCTGCCTCTTAGAGG
TGAAAATGAGAAACCCGATGTGACTATGGGGACTCCTGTTATGCCCACAGCAATGAGTTT
CCAAAACCTCTGCTGGCATGAACGGTGTGCCACAGCCATGGAATGAAAAAGAGGTTAAACG
AGAGAAGAGAAAACAGTCAAACCGAGAATCTGCTAGGAGGTCAAGACTGAGGAAGCAGGC
TGAAACAGAACAACTATCTGTCAAAGTTGACGCATTAGTAGCTGAGAACATGTCTCTGAG
GTCTAAACTAGGCCAGCTAAACAATGAGTCTGAGAAACTACGGCTGGAGAACGAAGCTAT
ATTGGATCAACTGAAAGCGCAAGCAACAGGGAAAACAGAGAACCTGATCTCTCGAGTTGA
TAAGAACAACCTCTGTATCAGGTAGCAAACTGTGCAGCATCAACTGTTAAATGCAAGTCC
GATAACCGATCCTGTGCGGGCTAGCTGACCGTGGCCGCAACATGAGAACCCGATATTTT
TTCCTTTGGGTTGTGATTGTAACCTAAAAGGAGACTTTTTGTTTTATTCTTAGATTGT
AGCTCTCTGCATAGTGAGCATAAATTGATGTAATATGGTTTAAGAGATTCCGGTGTCTCT
GGTGTGTGCTGCAACCACATAATTGGTGATAGATAGGTTTAGTTATATAAGCAAATGTAT
TAGAGATAAGGGGAGACATATTTGATGGTCTTT

>G561 Amino Acid Sequence (domain in AA coordinates: 248-308)
MGSNEEGNPTMNSDKPSQAAPEQSNVHVYHHDWAAMQAYYGPRVGIPQYYNSNLAPGHA
PPPYMWASPSMMAPYGAPYPFPCPGGVYAHPGVQMGSPQGPVSQSASGVTTPLTIDA
PANSAGNSDHGFMKKLKEFDGLAMSISNNKVGSAEHSSEHRSSQSSSENDGSSNGSDGNT
TGGEQSRKRQRQSPSTGERPSSQNSLPLRGENEKPDVTMGTPVMPTAMSFQNSAGMNG
VPQPNNEKEVKREKRQSNRESARRSRLRKAETEQLSVKVDALVAENMSLRSLKGLQNN
ESEKRLLENEAILDQLKAQATGKTENLISRVDKNNSVSGSKTVQHQLLNASPIITDPVAAS
*

>G590 (102..1223)
TCGACAGACACTCTCCCTCTCTCCATGCCCATAAATCTCAAAGACTGTTTTAAAAA
AATGTTTTAGCTTTAACTGCTTTTTTTTGTGTTGTTGTTGTAATGATATCACAGAGAGAAG
AAAGAGAAGAGAAGAAGCAGAGAGTGATGGGAGATAAGAAATTGATTTTCATCTTCTTCTT
CTTCTCGGTTTACGATACTCGTATCAATCATCATCTTCATCATCCTCCGTCTTCTTCCG
ACGAAATCTCTCAGTTTCTCCGGCATATTTTCGACCGTTCTTCTCCTTTACCTTCTTACT
ACTCCCCGGCGACGACTCAACGACGGCGTCTTTGATTGGTGTGCACGGGAGCGGTGACC
CACATGCAGATAACTCGAGAAGTCTCGTTTTCTCATCATCCACCGTCAGATTCTGTGCTTA
TGTCGAAACGTGTGCGAGATTTCTCTGAGGTTTTAATCGGCGGAGGATCAGGCTCAGCCG
CCGCGTGTTTTGGTTTCTCCGGTGGTGAATAATAACAACGTTCAAGGAAATAGCTCTG
GGACTCGAGTATCGTCTTCTCCGTTGGAGCTAGTGGCAACGAGACAGATGAGTATGACT
GTGAAAGCGAGGAAGGAGGAGAAGCTGTAGTTGATGAAGCTCCCTCTTCCAAGTCAGGTC
CTTCTTCTCGTAGTTTCATCTAAAAGATGCAGAGCTGCTGAAGTTTCATAATCTCTCTGAGA
AGAGGAGGAGAAGTAGAATTAATGAAAAATGAAAGCTTTACAAAGTCTCATCCCTAATT
CAATAAGACGGGATAAGGCTTCAATGCTTGATGAAGCCATTGAGTATCTGAAACAGCTTC
AGCTCCAAGTTTCAGATGTTGACTATGAGAAATGGAATAAACTTGCATCCTTTGTGTTTAC
CTGGAACCTACATTACCCCATGCAACTCTCTCAGATTTCGACCCCTGAAGCAACCAATG
ATCCTCTGCTTAATCATACCAATCAGTTTGCTTCGACTTCTAATGCACCGGAAATGATCA
ATACTGTGGCTTCTTCATACGCTTTGGAACCTTCTATTTCGAGTCACTTTGGACCTTTCC
CTCTCCTTACTTCACCCGTGGAGATGAGTCGGGAAGGTGGGTAACTCATCCAAGGTTGA
ACATTGGTCATTCCAACGCAACATAACCGGGGAACAAGCTCTGTTTGTATGGACAACCTG
ACCTAAAAGATCGAATTACTTGAACAGTGTCCCACTTCGGGATCTCTATGTGTTCTTGT
TTCTTAGAACGCAAGCCATAAAGCTGTCTGAC

>G590 Amino Acid Sequence (domain in AA coordinates: 202-254)
MISQREERBEKKQRYMGDKKLISSSSSSVYDTRINHHHLHHPSSSDEISQFLRHIFDRS
SPLPSYYSPATTTTTASLIGVHSGDPHADNSRSLVSHHPPSDSVLMSKRVGDFSEVLIG
GGSGSAAACFGFGGNNNNVQGNSSGTRVSSSVGASGNETDEYDCESEEGGEAVVDEA
PSSKSGPSSRSSSKRCRAAEVHNLSEKRRSRINEKMKALQSLIPNSNKTDKASMLDEAI
EYLKQLQLQVQMLTMRNGINLHPLCLPGTTLHPLQLSQIRPPEATNDPLLHNHTNQFASTS
NAPEMINTVASSYALEPSIRSHFGPFLLTSPVEMSREGGLTHPRLNIGHSNANITGEQA
LFDGQPDLDKDRIT*

>G818 (65..1060)
GTATTCTTACAATAAACGACCAAAAAGTTAATACAAGAAATAGAAACGGTGTAGGAAGC
TACTATGACGGCAATTCCAAACGTCGTCGATATTGAATCTTCTTCTCTTCTCGCTTTGTCA

AGAGACGGCAACGGAGACCGTCACCGTTGAAAGAGGCTCGTCTGATTTCATCTTCAAAGCC
AGACGACGTCGTTTTACTAATCAAGGAAGAGGAGGATGACGCCGTTAACTTGTCACTTGG
TTTTTGGAAATTGCACGAGATAGGTTTAATAACACCGTTCTTGAGAAAGACGTTTGAGAT
CGTCGATGACAAAGTAACAGACCCCGTTGTATCATGGAGCCCGACCCGTAAAAGCTTTAT
CATTTGGGATTCTTACGAGTTCTCAGAGAATCTACTTCCCAAATACTTCAAGCACAAGAA
CTTCTCCAGTTTTATTTCGTCAGCTTAACTCTTACGGTTTTAAAAAGGTCGATTTCAGATAG
GTGGGAATTTGCTAACGAAGGGTTTCAAGGAGGGAAGAAACATTTGCTTAAGAACATCAA
GAGGAGAAGCAAAACACTAAATGTTGTAACAAGGAAGCGAGTACCACCACGACAGAGAC
TGAGGTTGAGTCATTGAAGGAGGAACAGAGTCCAATGAGATTGGAGATGTTGAAGCTGAA
ACAACAACAAGAAGATCTCAACATCAGATGGTCACTGTGCAGGAGAAGATCCACGGAGT
TGATACCGAACACAGCATATGCTTAGTTTCTTGCAAAGTTGGCTAAAGATCAAAGATT
TGTAGAGAGACTGGTGAAGAAGAGAAAGATGAAAATACAGAGAGAGCTAGAAGCAGCTGA
ATTCTGTGAAGAAGCTCAAGTTGCTTCAGGATCAAGAACTCAAAAGAACTTGTAGATGT
AGAAAGAGAATTTATGGCCATGGCTGCAACAGAACACAATCCCGAGCCTGACATTTTGGT
GAACAATCAAAGCGGAATACGAGATGTCAGCTTAACTCAGAGGACCTACTTGTGACGG
TGGCTCAATGGATGTAAATGGGAGGATAGAGATAGAGTAGAGCAAAACCGGTAACATAGC
AATAGAGAAGGTACCAAATCCCAAGGCTTGAGATCCGAAT

>G818 Amino Acid Sequence (domain in AA coordinates: 70-162)

MTAIPNVVDIESSSSSLCQETATETVTVERGSSDSSSKPDDVLLIKEEDDAVNLSLGF
WKLHEIGLITPFLRKTFEIVDDKVTDPVVSWSPTRKSFIIWDSYEFSENLLPKYFKHKNF
SSFIRQLNSYGFKKVSDRWEFANEGFQGGKKHLLKNIKRRSKNTKCCNKEASTTTTETE
VESLKEEQSPMRLEMLKLKQQEESQHQMVTVQEKIHGVDTEQQHMLSFFAKLAKDQRFV
ERLVKKRKMKIQRELEAAEFVKKLKLQDQETQKNLLDVEREFMAMAATEHNPEPDILVN
NQSGNTRCQLNSEDLLVDGGSMDVNGRIEIE*

>G849 (218..2077)

AACTCGAGAATCTTTCATTTCTTTTAAATCTTAGAATCTCGAGTTTTTGTATAAATCGAT
TCTAATTTTTCTTTGTACATTGTTTTATATATACATAAAACACACAAATCGGGTATGGG
GGAATTTGGGTTTTAAGATAGCGTGATCTGTAATAATAAGTGGTTCGCGATCGTGATCAA
GAAACTGGTGGCTGATAGTGATATGCATATTTGAGAGATGGTGTTCAGAGAAAGTTAGA
TTGCCTTTCCGTGGGATTTGATTTTCCCAACATTCCCAGAGCTCCTCGTTCATGCAGGAG
GAAGGTTCTAAACAAGAGGATTGATCATGATGATAACACTCAGATCTGTGCAATTGA
CTTACTAGCTTTGGCTGGAAAGATTCTACAGGAAAGCGAGAGTTCCTCTGCGTCTTCTAA
TGCATTTGAAGAAATTAAGCAAGAGAAAGTAGAAAATTGCAAGACTATTAAATCTGAGTC
TTCTGACCAAGGAAACTCTGTGTCAAAGCCTACTTATGATATCTCTACTGAGAAGTGTGT
GGTGAACAGTTGTTTTTTCATTTCCGGATAGTGACGGCGTTTTGGAGCGGACTCCGATGTC
TGATTACAAGAAGATTTCATGGTTTGATGGATGTAGGGTGTGAAAACAAGAATGTAAATAA
TGGGTTTCGAGCAAGGAGAAGCAACCGATCGCGTGGGTGATGGAGGCTTAGTCACTGATAC
TTGCAACTTAGAGGATGCAACTGCGTTAGGTCTGCAGTTTCCGAAATCAGTCTGTGTGGG
TGGTGATTTAAATCACCATCCACCTTGGATATGACCCCTAATGGTTCCATGCTAGACA
TGGGAACCATACTAACCCTAGGTAGAAAAGATGATGATGAAAAATTCTATAGTTACCATAA
ACTTAGCAATAAATTTAAGTCGTATAGGTCTCCAACAATTGCAAGAATAAGAAAGTCCAT
GTCGTCCAAATACTGGAACAAGTTCCAAAAGATTTTGGATAACAGTAGAGCTGATGTGGG
TGTGAAGACTCTTTATCGCAAAAGAAAATCATGTTATGGTTACAACGCATGGCAGCGTGA
GATCATTTATAAGAGAAGAAGATCACCTGACAGAAGCTCGGTGTAACCTTCTGATGGAGG
ACTCAGTAGTGGAAGTGTTTCCAAGTTACCCAAGAAGGGAGATACAGTAAAGCTAAGCAT
TAAGTCCTTTAGGATTCAGAGCTTTTTATTGAAGTTCCAGAAACTGCAACAGTAGGATC
ACTAAAGAGGACTGTGATGGAGGCTGTCTAGTGTCTTACTCAGCGGAGGAATACGTGTTGG
GGTGTAAATGCATGGGAAGAAGGTTAGAGATGAAAGGAAAACCTCTGTCCAGACTGGGAT
CTCATGTGATGAAAATCTAGACAACCTTGGGTTACCTTGGAGCCTAGTCCCAGCAAAGT
TCCCCTACCTTTGTGTTCTGAAGATCCTGCTGTGCCAACCGACCTACAAGTTTGTCTGA
ACGGTCTGCGGCGTCTCCTATGCTAGATTCTGGAATTCACATGCAGATGACGTGATTGA
TTCAAGAAATATTGTGGACAGTAACCTCGAATTAGTTCCATATCAGGGTGACATATCTGT
TGATGAACCTTCATCAGATTCAAAGAGCTTGTCCCACTTCAGAGTTGGAAGTCAAGGC
GCTTGCCATAGTTCCGTTGAACCAGAAACCTAAGCGTACTGAGCTAGCCCAGAGGAGAAC
TAGGAGACCCTTCTCTGTGACAGAGGTAGAAGCTTGTGACAAGCAGTTGAGGAACCTCGG
GACTGGAAGATGGCGTGATGTAAATGCGTGCTTTGAGGATGCAGATCATCGGACTTA
CGTGGAATTGAAGGACAAATGGAAGACGCTAGTTCACACAGCAAGTATATCCCCACAGCA

ACGAAGAGGAGAGCCGGTGCCACAAGAACTGCTAGACAGAGTCTTGAGGGCATAACGGGTA
TTGGTTCGCAGACCAAGGAAAACATCAGGCGAGAGGAGCGTCCAAAGATCCAGACATGAA
CAGAGGTGGAGCTTTTGAATCAGGTGTTTCAGTGTAAGGAGGAGGTACGCATTGGTGGG
TGGGTGTACAGAAGCAAACAACACAATAAATGGACAACCTCAATTTCTGCAAAGTTTAATT
GTCTTTATTTCTCGTTTTTTTTTTTTTCTCTCTACATACACTTTTTTTTTTCT
>G849 Amino Acid Sequence (domain in AA coordinates: 324-413, 504-583)
MVFKRLDCLSVGFDFPNIPRAPRSCRRKVLNKRIDHDDNTQICAIDLLALAGKILQES
ESSSASSNAFEEIKQEKVENCKTIKSESSDQGNSVSKPTYDISTEKCVVNSCFSFPDSG
VLERTPMSDYKKIHGLMDVGCENKNVNNNGFEQGEATDRVGDGGLVTDTCNLEDATALGLQ
FPKSVCVGGDLKSPSTLDMTPNGSYARHGNHTNLGRKDDDEKFYSYHKLSNKFYSYRSP
IRRIKSMSSKYWKQVPKDFGYSRADVGVKTLYRKRKSCYGYNAWQRELIYKRRSPDRS
SVVTSDDGLSSGSVSKLPKKGDTVKLSIKSFRIPELFIEVPETATVGSILKRTVMEAVSVL
LSGGIRVGVLMHGKKVDERKTLSTGTISCDENLDNLGFTLEPSPSKVPLPLCSEDPAVP
TDPSTLSERSAASPMLDGIPHADDVIDSRNIVDSNLELVPIYQGDIVDEPSSDSKELVP
LPELEVKALAIIVPLNQPKRTELAQRRTRRPFSVTEVEALVQAVEELGTGRWRDVKLRAF
EDADHRTYVDLKDWKTLVHTASISPQQRGEPVPQELLDRVLRAYGYWSQHQQKHQARG
ASKDPDMNRGGAFESGVSV*

>G892 (21..1004)
TATAACAATTCTTCCAACAATGTCAATTGAGTCAGCCAATAACACGGACCGATAGTGCAC
CCAATGGAGCATTTAGGACTTTTGGTCTCTACTGGTGTACCATTGTGATCGTATGGTCA
GAATTGCATCCTCTAACCCATCAGAGATCGCCTGTCTCGATGTTTGAGGCAATTTGTGCG
TTGAGATTGAAACGAGACAACGGCCTCGGTTTACTTTCAACCATGCTACTCCGCTTTTG
ATGCTTCTCTGAGGCTCGTCTTCTCGAAGCTCTCTCGCTCATGTTTGAGCCTGCAACCA
TAGGTAGGTTTGGTGCAGACCCATTTCTTAGGGCAAGATCCAGAAACATCTTGAACCTG
AATCAAGACCCCGACCGCAACATCGAAGACGACACAGCCTTGACAATGTTAACAATGGTG
GTTTACCTCTACCAAGAAGAACATATGTTATTCTCCGGCCCAATAATCCGACTAGTCCAC
TCGGAACATAAATTCGCCACCACCAATCAAGCACCACCACGGCATGTGAACCTCACATGATT
ACTTTACTGGAGCATCAAGCTTAGAGCAGCTGATTGAACAACCTAACACAAGACGATAGGC
CTGGACCACCACCTGCGTCAGAACCCACCATTAAATCCCTACCATCTGTGAAAATAACAC
CACAACATCTAACTAACGACATGTCCCAATGCACAGTGTGCATGGAAGAATTCATTGTTG
GTGGGGACGCAACGGAATTACCATGTAAACATATTTACCATAAAGATTGTATAGTCCCGT
GGCTTAGGCTTAAACAATTCTTGCCCTATCTGCCGCCGTGACCTGCCACTTGTCAACACCG
TTGCTGAATCTCGAGAAAGGAGCAATCCTATTAGACAAGACATGCCTGAAAGAAGGCGTC
CAAGGTGGATGCAACTCGGTAACATTTGGCCATTTAGAGCAAGATAACCAAGGGTTAGTC
CAGAAGAAACAGCAAACAGAAATCCTCGAGATAACAGGAGCTAACTCTGAATATTCATG
GGAAATAAAAATCGTGACTATCTATATGTATAGACTCTATGAGACATTGTCTATTTGAAT
GTGCATGTATATCTCAGAAATAAATCAAGCGAAACATATTTAACGACTAAAAAAA

>G892 Amino Acid Sequence (domain in AA coordinates: 177-270)
MSLSQPITRTDSAPNGAFRTFGLYWCYHCDRMVRIASSNPSEIACPRCLRQFVVEIETRO
RPRFTFNHATPPFDASPEARLLEALSLMFEPATIGRFADPFLRARSNILEPESRPRPQ
HRRRHSLDNVNNGGLPLPRRTYVILRPNNPTSPLGNI IAPPNQAPPRHVNSHDYFTGASS
LEQLIEQLTQDDRPGPPASEPTINSLPSVKITPQHLTNDMSQCTVCMEEFIVGGDATEL
PCKHIYHKDCIVPWLRLNNSCPICRRDLPLVNTVAESRERSNP I RQDMPERRRPRWMQLG
NIWPFRRARYQRVSP EETANQNPRDNRS*

>G961 (1..1200)
ATGTCAAAATCTATGAGCATATCAGTGAACGGACAATCTCAAGTGCCTCCTGGGTTTAGG
TTTCATCCGACCGAGGAAGAGCTGTTGCAGTATTATCTCCGGAAGAAAGTTAATAGCATC
GAGATCGATCTTGATGTCAATTCGCGACGTTGATCTCAACAAGCTCGAGCCTTGGGACATT
CAAGAGATGTGTAATAAGGAACAACGCCACAAAACGACTGGTATTTCTTTAGCCACAAG
GACAAAAATATCCGACGGGAACGAGAACTAACAGAGCCACTGCGGCTGGATTTTGAAA
GCAACTGGCCGCGACAAGATCATATATAGCAATGGCCGTAGAATTGGGATGAGAAAGACT
CTTGTTTTCTACAAAGGCCGAGCTCCTCACGGCCAAAAATCTGATTGGATCATGCATGAA
TATAGACTCGATGACAACATTATTTCCCCCGAGGATGTACCGTTCATGAGGTCTGTGAGT
ATTATAGGGGAAGCATCACAAGACGAAGGATGGGTGGTGTGTCGTATTTTCAAGAAGAAG
AATCTTCACAAAACCTAAACAGTCCCGTCGGAGGAGCTTCCCTGAGCGGCGCGGAGAT
ACGCCGAAGACGACATCATCTCAGATCTTCAACGAGGATACTCTCGACCAATTTCTTGAA
CTTATGGGGAGATCTTGTAAGAAGAGCTAAATCTTGACCCTTTCATGAACTCCCAAAC

CTCGAAAGCCCTAACAGTCAGGCAATCAACAACCTGCCACGTAAGCTCTCCCGACACTAAT
CATAATATCCACGTGAGCAACGTGGTCGACACTAGCTTTGTTACTAGCTGGGCGGCTTTA
GACCGCCTCGTGGCCTCGCAGCTTAACGGACCCACATCATATTCAATTACAGCCGTCAAT
GAGAGCCACGTGGGCCATGATCATCTCGCTTTGCCTTCCGTCCGATCTCCGTACCCAGC
CTAAACCGGTCCGCTTCGTACCACGCCGTTTAAACACAGGAATATACACCGGAGATGGAG
CTATGGAATACGACGACGTCTCTATCGTCATCGCCTGGCCCATTTTGTACGTGTCTG
AATGTTTTGCTGCTTTGTTGTCTCTCTTCGTCTGCAGCTTCAGTTCTGGCCGTTCCAACCA
TGGCAGAGGCAGGTTTCATTTTCATCTTCATCGCCTCAGATGCAGATCTCTCTCCATTGA
>G961 Amino Acid Sequence (conserved domain in AA coordinates: 15-140)
MSKSMSISVNGSQVPPGFRFHPTBEEELLQYYLRKKVNSIEDLDVIRDVDLNKLEPWDI
QEMCKIGITTPQNDWYFFSHKDKKYPTGTRTNRATAAGFWKATGRDKIIYSNGRRIGMRKT
LVFYKGRAPHGQKSDWIMHEYRLDDNIISPEDVTVHEVVSIIGEASQDEGWVVCRIFFKKK
NLHKTLSNPVGGASLSGGGDTPKTTSSQIFNEDTLDQFLELMGRSCKEELNLDPFMKLPN
LESPNSQAINNCHVSSPDNHNHVSINVVDTSFVTSWAALDRLVASQLNGPTSYSITAVN
ESHVGHDLALPSVRSYPYPSLNRASYSYHAGLTQEYTPEMELWNTTSSLSSSPGPFCHVS
NVLLLVCLLRLQLQFWFPQWQVHFDLSSPQMQLSLH*

>G1465 (163..1125)
TATCCTTCGCAAGACCCTTCCTCTATATAAGGAAGTTCATTTTCATTTGGAGAGGACACGC
TGACAAGCTGACTCTAGCTTATCTGGTACCGTCGACCTCATTCTTGCGTTTGATCTTTCT
TTCTCTAGATCCCATATTTTTCTTGATCAATTTAGTTTCATTATGGAGGAAGATGCAGCT
TTTGATCTACTCAAAGCCGAACCTCTTAAACGCAGAAGACGATGCAATAATCTCACGTTAT
CTGAAGCGTATGGTTCGTCACCGGAGACTCATGGCCTGATCATTTCATCGAAGACGCAGAC
GTGTTCAACAAGAATCCAAATGTGGAGTTCGATGCTGAGAGCCCTAGCTTCGTGATAGTT
AAACCTCGAACAGAGGCTTGTGGTAAACCGATGGATGTGAAACTGGTTGCTGGAGGATC
ATGGGTCTGTGATAAAACCGATAAAATCGACGGAGACTGTGAAGATTCAAGGGTTCAAGAAG
ATTCTCAAGTTCCTGCCTAAAGAGGAAACCTAGAGGATACAAGAGAAGTTGGGTAATGGAA
GAGTATAGGCTTACCAATAACTTGAAGTGAAGCAAGATCATGTGATTGCAAGATTCGG
TTTATGTTTGAAGCTGAAATCAGTTTCTTGCTAGCCAAGCATTCTTACACTACATCAGAA
TCACTTCCTCGAAATGAGCTGTTGCCAGCTTACGGATTCTTTTCATCAGATAAGCAATTG
GAGGATGTATCTTATCCGGTGACGATAATGACTTCTGAAGGAAACGATTGGCCTAGCTAC
GTTACCAACAATGTGTATTGTCTGCATCCATTGGAGCTCGTTGATCTTCAAGATCGGATG
TTTAATGATTACGGAACCTGCATCTTCGCTAACAGACTTGTGGTAAACCGATAGATGC
ATTAATGGTGGTTTACTGGAAAATTTTGCACCGTGATAGGCTGATCAAGTCAAAGTCCGGG
ATAGTTATTGGTTTCAAGAAGGTGTTTAAGTTTCATGAAACGGAGAAAGAAAGATACTTC
TGTGGTGGAGAAGATGTGAAGGTAACCTGGACTCTAGAAGAGTATAGGCTTAGCGTGAAG
CAGAATAAATCTTGTGCGTTATCAAGTTTACTTATGATAACTAAGAATCTTTTCTTTGG
ATTTTATGATCATCTTAGTATCGCGACCGCTCTAGACAGGCCTCGTACCGGATCCTCTAG
CTAGAGCTTTTCGTTTCGTATCATCGGTTTCGACAACG

>G1465 Amino Acid Sequence (conserved domain in AA coordinates: 242-306)
MEEDAAFDLLKAELLNAEDDAIISRYLKRNVNGDSWPDHFIEDADVFNKNPNVEFDAES
PSFVIVKPRTEACGKTDGCETGCRIMGRDKPIKSTETVKIQGFKKILKFKLKRKPRGYK
RSWVMEYRLTNNLNWKQDHVICKIRFMFEAEISFLAKHFYTTSESILPRNELLPAYGFL
SSDKQLEDVSYPTVIMTSEGNDWPSYVTNNVYCLHPLELVDLQDRMFNDYGTICIFANKTC
GKTDRCINGGYWKILHRDLIKSKSGIVIGFKKVKFPHETEKERYFCGGEDVKVWTWLEE
YRLSVKQNKFLCVIKFTYDN*

>G425 (45..1196)
GAAAACAGTCTTCTCTTCTCCGATCCCAAAAACGCAGGAAAACAATGTGCTTTAACAGCTCCC
ACCTCCTTCCTCCAACAAGAAGACCTTCCTCTCCGACACTTCACCGATCAATCACAGCAACCTC
CGCCGACGCGTCACTTCTCTGAAACACCTTCGCTTGTACCGCCAGTTTCCTCAACCTCCCTA
CCACCCTTACCATTGCGGATCCGATCTCGCTCCTCCGACCGCAACGGAGACAATTCCGTT
GCTGATACAAACCCACGCTGGCTCTCCTTTTCATTTCGGAGATGCAAAATACTGGAGAAGTACG
TTCTGAAAGTTATCGACGGAGTCAACGCCGATGGTGAAACGATACTCGGCGTTGTAGGAGGT
GAAGATTGGCGGAGTGCTAGCTATAAGGCGGCGATTTTAAAGACATCCGATGTACGAGCAGC
TTCTTGGCGCTCACGTGGCTTGCCTTAGGGTTGCGACTCCCGTTGACCAGATTCCGAGGATC
GATGCTCAGCTCAGTCAGTTGCATACCGTCGCCGCGAAATACTCCACTCTTGGTGTGGTTGTT
GACAACAAGGAACCTTGATCATTTTCATGTACATTATGTTGTCTTGTTATGTTTCATTTAAAGAACA
ACTCCAACACCACGTTTGTGTCCATGCAATGGAAGCCATTACCGCTTGTGGGAGATTGAACA

>G2069 (1..1026)

ATGGAAGGAGGAGGAAGAGGACCAAATCAAACGATTCTCAGTGAAATAGAACATATGCCT
GAAGCTCCACGTCACGTATCTCTCATCACCCTCGAGCTCGCTCTGAAACCTTCTTCTCC
GGCGAATCAATCGACGATCTCCTCTTATTCGATCCTTCCGATATCGATTCTCTTCTCTA
GACTTCTCAACGCTCCACCACCACCACAACAATCACAACAACAACCGCAAGCTTCTCCC
ATGTCCGTTGATTCCGAAGAAACCTCATCGAACGGTGTGTCTCTCTAATTCTCTTCTCT
CCAAAACCCGAAGCTAGATTTCGGTCCCATGTTCTGATCTTCTCGGTTGATTCCGATTTC
TTCGATGATTTGGGTGTTACTGAGGAGAAGTTTATAGCTACAAGTTCAGGAGAGAAGAAG
AAAGGGAAATCATCATCATAGCAGGAGTAATTCTATGGATGGAGAGATGAGTTCCGGCGTCG
TTTAATATCGAATCGATTTTAGCTTCTGTGAGTGGTAAAGATAGTGGGAAGAAGAATATG
GGTATGGGTGGTGATAGACTTGCTGAGCTTGCTTTGCTTGATCCTAAAAGAGCTAAAAGG
ATTTTAGCGAATAGACAATCTGCGGCGAGGTCGAAAGAGAGGAAGATTAGGTATACCTGGT
GAGTTAGAGAGGAAGTTTACAGACTTTCAGAAATGAAGCTACTACATTGTCTGCTCAAGTC
ACTATGTTAGCAGAGAGGAACATCAGAGCTGAACACTGAAAATAAACACCTCAAAATGCGG
CTTCAAGCTTTAGAGCAACAAGCTGAACCTAGGGATGCTTTGAATGAAGCGCTGCGGGAT
GAACTGAACCGACTTAAGGTGGTAGCTGGAGAAATTCCTCAGGGGAATGGAATTCCTTAC
AACCCTGCTCAATTCTCATCTCAGCAATCGGCAATGAATCAGTTTGGGAACAAAACGAAC
CAACAGATGAGTACAAACGGGCAGCCATCGCTCCCAAGCTACATGGATTTCACCAAGAGA
GGCTGA

>G2069 Amino Acid Sequence (domain in AA coordinates: TBD)

MEGGGRGPNQILSEIEHMPAPRQRISHHRRARSETFFSGESIDDLILFDPSDIDFSSL
DFLNAPPPPPQSQSQPQASPMVDSEETSSNGVVPNSLPKPEARFGRHVRFSFVDSDF
FDDLGVTEEFKFIATSSGEKKGNHHHSRSNSMDGEMSSASFNIESILASVSGKDSGKKNM
GMGGDLAELALLDPKRAKRIILANRQSAARSKERKIRYTGELERKVQTLQNEATTLQAQV
TMLQRGTSELNNTENKHLKMRQLALEQQAELRDALNEALRDELNLKVVAGEIPQGNNGSY
NRAQFSSQQSAMNQFNGKTNQQMSTNGQPSLPSYMDFTKRG*

>G1852 (55..1857)

CATCTGATCTGCTCTCGAAGACGAAAGCTTCGAGTACTGGTTGAAGCTAAAGCTATGGGA
CACGTGAATCTACCTGCATCAAAGCGTGGTAACCTCGTCAATGGCGTCTCCTCGACATC
GTAACCGCTGCTTCTCTCGGTATCGTACTTCTCTTCTTCACTCTTTTATTCACCTCTT
GGTGATTCCATGGCGGCTTCTGGTCCGCAACGCTGCTTCTCTTACGGCGTCAGATCCG
AGGCAACGGCAGCGATTAGTGACTTTGGTTGAAGCTGGTCAGCATTGCAACCGATCGAG
TATTGTCTCTGCGGAAGCTGTTGCTCATATGCCTTGTGAGGATCCGAGAAGGAATAGTCAG
CTTAGTAGAGAGATGAATTTCTATAGGGAGAGACATTGTCTTTGCTTGAGGAGACTCCG
CTCTGTTTGATTCTCTCGCCTTCTGGTTATAAAATTCCTGTTCGGTGGCCTGAGAGTCTT
CACAAGATTTGGCATGCAAAATGCCATATAACAAAATTCCTGACCGGAAAGGTATCAA
GGATGGATGAAAAGGGAAGGGGAATACTTTACTTTCCAGGCGGTGGCAGATGTTTCTCT
GGCGGAGCTGGCCCAATACATTGAAAAGCTTGACAGTATATTCGGCTTAATGGTGGAACT
TTGAGAACTGCTCTTGACATGGGATGCGGGGTAGCTAGTTTGGAGGTACTCTACTATCT
CAAGGCATTCTAGCCCTCTCAATTGCTCCAAGAGATTACATAAATCTCAAATTCAGTTC
GCTTTGGAAAGAGGAGTGCCCTGCATTTGTTGCCATGCTTGGCACTCGTAGACTCCCTTT
CCTGCATACTCTTTGACCTGATGCACTGTTCCCGATGTTTGATTCTTTTACGGCTTAC
AATGCAACTTACTTCATCGAAGTAGATAGGTTACTGCGCCCTGGAGGATATCTTGTAATC
TCTGGCCACCTGTACAATGGCCATAACAAAGACAAAGAATGGGCTGATCTTCAGGCGGTG
GCTAGAGCTTTGTGCTATGAGCTAATTGCGGTTGATGGAAACACTGTCATCTGGAAGAAG
CCTGTTGGAGATTCTATGTCTACCTAGCCAGAATGAGTTTGGGCTTGAGTTGTGTGATGAG
TCTGTTCCGCAAGTGATGCATGGTATTTTAAATTGAAGAGGTGTGTTACCAGGCCATCA
TCCGTCAAAGGAGATCACGCTTTGGGAACATATCCAAGTGGCCGGAGAGGCTTACTAAA
GTTCTCTCTAGGGCCATTGTCTGAAAAACGGATTGGATGTGTTTGAAGCAGATGCAAGG
CGGTGGGCAAGACGCGTTGCTTATTACAGGGATTCTCTTAAGCTGAAATCTCCA
ACTGTCCGCAATGTCTAGGACATGAACGCAATCTTCCGAGGCTTTGACAGCAACCCCTGCA
TCTGATCCTGTGTGGGTTATGAATGTCAATCCAGCTCGGAAGCCATTAACTCTTGACGTG
ATTTATGACAGAGGTCTCATCGGTGTTTACCATGATTGGTGTGAACCATTTTCAACATAT
CCCCGCACGTATGATTTTCATCCATGTATCAGGAATTGAATCACTGATAAAACGACAAGAC
TCAAGCAAATCGAGGTGTAGCTTAGATCTAATGGTAGAGATGGACAGAATATTACGT
CCAGAAAGGAAAGGTTGTGATCCGAGACTCTCTGAGGTGCTAGATAAAGTCGCACGAATG
GCTCATGCTGTAAGATGGTCTTCTTCCATACACGAGAAAGAACCTGAATCCCATGGAAGA

YDMPSSDGTGGYSGWTSESQGSNPGGVFTMWNE*

>G761 (521..1549)

GGGGCCGACCGGCCCGCCGGGCAGGTCTAGGTTCAAAAGGACTCACAAGAGAGAGATAGT
 ATGATTGATAGGGAAAGAGAGAGATGAAAGAAAGTAAATATATAATAGATTATTAGG
 ACACGAGTGTCTCTTTTGTGTTTGTGTCTTGTGTCTCTCTCTTTCTCTCTCCTCGAA
 TGATCATCTTTATATAACCCCTACTCTCTTTCTCTTTTCCATTCTTTTCATATCATTCTCC
 CTTTCTCTCTCGGGATCTGATCTCTCTTTCCAGTAACCTATTCCCGAGGAGCACTGTCAA
 ATCTTGTCCACTCTTTGATCTTATCTCGATCTCTTTCTCTTTCTAGTCTTGTGTAGTCTT
 CAACTTGTGATGTTATCTATATAGTAATCACGAGAGAGAATCATACAATAGCTGAAACA
 TAAAGCTTTCTTAGAAGCTTTAAAAAGGTCTCATCTGGATTATCCTGTTTAAATTTCTAGA
 GTTTCTTCAGGCAGATTATTAACCGATCAAGAAGACAAAATGAATTCATTTTCCCACGT
 CCCTCCGGGTTTTAGATTTCACCCGACAGATGAAGAACTTGTAGACTACTACCTGAGGAA
 AAAAGTCGCATCGAAGAGAATAGAAATTGATTTTATAAAGGACATTGATCTTTACAAGAT
 TGAGCCATGGGACCTTCAAGAGTTGTGCAAAATTGGGCATGAAGAGCAGAGTGATTGGTA
 CTTCTTTAGCCATAAAGACAAGAAGTATCCACAGGGACTCGAACCAATAGAGCAACAAA
 AGCAGGGTTTTGGAAGCCACCGGAAGAGATAAGGCTATCTATTTGAGGCATAGTCTAAT
 TGGCATGAGGAAAACTTGTGTTTTACAAGGGAAGAGCCCCAAATGGACAAAAGTCTGA
 TTGGATCATGCACGAATACCGCTTAGAAACCGATGAAAACGGAACCTCCTCAGGAAGAAGG
 ATGGGTTGTGTGTAGGGTTTTCAAGAAGAGATTGGCTGCAGTTAGACGAATGGGAGATTA
 CGACTCATCCCTTCACATTGGTACGATGATCAACTTTCTTTTATGGCCTCCGAGCTCGA
 GACAAACGGTCAACGACGGATTCTCCCAATCATCATCAGCAGCAGCAGCAGCAGCAGCA
 ACAACATATGCCATATGGCCTCAATGCATCTGCTTACGCTCTCAACAACCTTAATTTGCA
 ATGCAAGCAAGAGCTAGAACTACACTACAACCACCTGCAATCAAATATCGCGCATGAGGA
 ACAATTGAATCAAGGAAATCAGAACCTCAGCTCTCTATACATGAACAGCGGCAACGAGCA
 AGTGATGGACCAAGTCAAGACTGGAGAGTTCTCGATAAATTTGTTGCTTCTCAGCTAAG
 CAACGAGGAGGCTGCCACAGCTTCTGCATCTATACAGAATAATGCCAAGGACACAAGCAA
 TGCTGAGTACCAAGTTGATGAAGAAAAAGATCCGAAAAGGGCTTCAGACATGGGAGAAGA
 ATATACTGCTTCTACTTCTTCGAGTTGTGAGATTGATCTATGGAAGTGAGCTGAAAGAGA
 AGACATATAAATGCATATATACATATATATATATACGTACACAGCAACTAATCAAGTG
 TAGATGATGATGATGGTACAGATTTATATTTGCTTTGATTGATTCTTACTACATTATTGA
 ACTTATGTCATATGCATATATACATTTGCGTATCTATGCATATTTATACTTGTACTCAATA
 TGATTAACCATATATAAATCTAATCTAAATGTAACCTCCAATATTTTTTAAATAGACAAT
 TGTCTCTTCTTATTAGAAAAAAA

>G761 Amino Acid Sequence (domain in AA coordinates: 10-156)

MNSFSHVPPGFRFHPTDEELVDYLRKKVASKRIEIDFIKDIDLYKIEPWLQELCKIGH
 EEQSDWYFFSHKDKKYPGTTRTNRAKAGFWKATGRDKAIYLRHSLIGMRKTLVFKGRA
 PNGQKSDWIMHEYRLFTDENGTPQEEGWVVCRVFKRLAAVRRMGDYDSSPSHWYDDQLS
 FMASELETNGQRRILPNHHQQQHEHQHMPYGLNASAYALNNPNLQCKQELHLYNHLQ
 SNIAHEEQNLQGNQNFSSLYMNSGNEQVMDQVTDWRVLDKFVASQLSNEEAATASASIQN
 NAKDTSNAEYQVDEEKDPKRASDMGEEYTASTSSSCQIDLWK*

>G1056 (10..798)

GCTACATATATGGGTCTTATTAGAGGAAACATTGAAGAGCCTATATCTCAGTCATTAACG
 AGGCAGAACTCTCTCTATAGCTTAAAGCTCCATGAGGTTCAAACCCACTTAGGAAGTTCT
 GGAAAACCACTAGGAAGCATGAACCTTGATGAGCTTCTCAAGACTGTCTTGCCACCAGCT
 GAGGAAGGGCTTGTTTCGTGAGGGAAGCTTGACGTTACCTCGAGATCTCAGTAAAAAGACA
 GTTGATGAGGTCTGGAGAGATATCCAACAGGACAAGAATGGAACGGTACTAGTACTACT
 ACTACTCATAAGCAGCCTACACTCGGTGAAATAACACTTGAGGATTTGTTGTTGAGAGCT
 GGTGTAGTGACTGAGACAGTAGTCCCTCAAGAAAATGTTGTTAACATAGCTTCAAATGGG
 CAATGGGTTGAGTATCATCATCAGCCTCAACAACAAGGGTTTATGACATATCCGGTT
 TCGAGATGCAAGATATGGTGATGATGGGTGGATTATCGGATACACCACAAGCGCCTGGG
 AGGAAAAGAGTAGCTGGAGAGATTGTGGAGAAGACTGTTGAGAGGAGACAGAAGAGGATG
 ATCAAGAACAGAGAATCTGCAGCAGCTTACGAGCTAGGAAACAGGCTTATACACATGAA
 TTAGAGATCAAGGTTTCAAGTTAGAAGAAGAAAACGAAAACTTCGGAGGCTAAAGGAG
 GTGGAGAAGATCCTACCAAGTGAACCACCACAGATCCTAAGTGAAGCTCCGGCGAACA
 AACTCTGCTTCTCTCTGATCCTAAAGACTCTTCTTTCTTTCTTCTTTGTTGTTGGTTT
 ATATCAGACCGCTTTGTTCTTTGTATATTGTGTAGACTTTATTGACTTTGAACAGCATGT
 CTTTATAAACATTTCTTGAGTGT

>G1056 Amino Acid Sequence (domain in AA coordinates 183-246)
 MGSIRGNIEEPISQSLTRQNSLYSLKLHEVQTHLGSSGKPLGSMNLDELLKTVLPPEEG
 LVRQGSLLTPRDLSSKKTVDVWRDIQQDKNGNGTSTTTTHKQPTLGEITLEDLLLRAGVV
 TETVVPQENVVNIASNGQWVEYHHQPPQQQGFMTYPVCEMQDMVMMGGLSDTPQAPGRKR
 VAGEIVEKTVERRQKRMIKNRESAARSARKQAYTHELEIKVSRLEBENEKLRLRLKEVEK
 ILPSEPPDPKWKLRRTNSASL*

>G1447 (82..1086)

AAAAACCTAACCTAATTCTCTCAAGACAACCTCAAAGGTCTCTCCTTTTTTAGGTTTAT
 TATCACTTCCGTATAATCGCCATGTCTTCTTACCATGGAAAAACCAAAATCGAGTCGA
 ATCTTAAGATTCAATTTCTGAGTTTCAACAATCACCGTTCGTTGAAACTGGCTTTCCAACT
 TCTCTGATCGATCTCTTCTTCAAGAATCGCGATCGTCTAAAAAATCTCCATCTAAACGC
 TTCCAACGAATCGAACGCCAGATTCTGAACCGCTCCAAACGCTTCTTCGTTGAGTAATCAA
 GATACGATTTTGAAGAGCCCTCGAGGATTAAACCGTTTGAAGTAAGGTCGAGAAAGTT
 AATTGCGTTAAAGGTAAATCAGCGGCGTGAAGAAGACGCGATTAAAAATAGCGTTTTC
 GGCGGTAGCGGTGAGGTGCTTTTGTATGGCGTTTAAAGGTTTGTATAGTAGCGTTGCTCGCC
 TTGAGCACGAAGAAGAAGCTCACTTTAGGAATCACTCTCTGCTTTCGCTCTTCTCTTA
 ACAGAGCTCGTGGCGGCGCGTGTTCACGCGCTCTAATAACACCGACAAAGACAAAAAC
 GCGATTGCCCGCGAGAAAAATCGAACTTTTGTATGAACTCGAGTTCCCAAGCGATTCCA
 TGTCTGAGGAAACAGAGCATGTAGTATCTGAAACAGAGGTTTGAAGTTGAAAGGTTTA
 ACGATACGTGATCTGTGTCAAAGGACGAGAAATCAACAAGTAAAGTTGGAGACTAAAA
 TCGAAGATTGTGAAGAAGTTGAGGAGTTACAATAAGAAGGATAAGAAGACGATGAAGATC
 AAAGAAGAGTCTTTGATTGAAGTCTCGAGTTTGGTTTTAGAAGATAAACCAAGAAAATT
 GAGTCTGAGAGAGACGAAGAAGAAACGTTGAATCCTCCAGTGGTTGGATCAAACTGAAT
 GGGATTGTTCTGATCGTGATTGTGCTAACCGGTTTGTATGTGGGAAGGTCTTAGCTATT
 GTTCTGACACTATCATGTTTGGTTCTTAGATTAGGAGCAGTCAAAAAAGTTAATCTTTGC
 ATATAATTTTTTTTTGTTTTTTTAAACATGCTTGCATGTGAACTGTAAATTTTTCTCATT
 CATATGAAGGAGATTGGATTGAATGTTGAATACTAAA

>G1447 Amino Acid Sequence (domain in AA coordinates: 3-54, 124-156)

MSSLPWKKPKSSRILRFISEFQQSPFVETGFPTSLIDLFFKNRDLKKSPSKRFQRIERO
 IRTAPNASSLSNQDTIFEKPSRIKTVRSKVEKVNVCVKGSAALKKNNAIKNSVFGGSGEVV
 LMAFKVLIVALLALSTKKKLTGLITLSAFALLLTELVAARVFTSRNNTDKDNIAIREKI
 ETFDETRVPAIPCEPTEHVSETEVSKLGLTIRDLLSKDEKSTSKSWRLKSKIVKKL
 RSYNKKDKMTMKIEESLIEVSSLVLEDKPKKIESERDEEETLNPVVGSNLNGIVLIVI
 VLTGLLCGKVLAIVLTLSCVLRLGAVKKVNLCI*

>G323 (77..826)

CTGCTCATATCAGCCATTGACACAGTTGCTTTGGGTTTCCCTCAAACGGCGCCGATTGTC
 TGGATTTTGACCACTGATGGCCTTAGATCAATCTTTTGAAGATGCTGCTTTACTTGGAGA
 ACTCTATGGAGAAGGTGCATTTTGTTCAGAGCAAGAAACCTGAACCCATTACAGTCTC
 GGTTCCTTCTGATGATACTGATGATTGCAATTTTACTGCAATATTTGCTTAGACTCGGT
 GCAAGAACCTGTTGTGACTCTCTGTGGTCACCTCTTTTGTGCTGGCCTTGTATTCACAAATG
 GCTTGATGTACAGAGCTTCTCAACAAGTGATGAATACCAAGACATAGACAGTGTCTCTGT
 TTGTAATCTAAAGTTTCTCATTCTACTTTGGTTCTTTTGTATGGTAGAGGCCGTTGTAC
 TACTCAGGAGGAAGGTAAAAACAGTGTGCCTAAAAGACCCGTAGGACCGGTTTATCGGCT
 TGAAATGCCGAATTCACCTTATGCAAGTACTGATCTGCGGTTATCACAACGGGTTTCATT
 CAATAGCCACAGGAAGGTTACTACCTGTCTCAGGGGTGATGAGCTCGAACAGTTTATC
 ATACTCTGCTGTTTTGGATCCGGTGATGGTGATGGTTGGAGAAATGGTAGCTACGAGGTT
 GTTTGGAACACGAGTGATGGATAGATTTGCGTATCCGGACACTTACAATCTCGCAGGGAC
 TAGCGGGCCGAGGATGAGAAGGCGGATAATGCAGGCAGATAAATCGCTGGGAAGAATCTT
 CTTCTTCTTTATGTGTTGTGTTGTTCTGTGTTCTCTTCTTGTGTTTAGGTTTTCATAGCTAG
 CTTGGTTCTGCTACTGTTTCAGTTTCTTCAGG

>G323 Amino Acid Sequence (conserved domain in AA coordinates: 48-96)

MALDQSFEDAALLGELYGEGAFCKSKKPEPITVSPDDTDDSNFDCNICLDSVQEPVV
 TLGHLFCWPCIHKWLVDVQSFSTSEYQRHRQCPVCKSKVSHSTLVPLYGRGRCTTQEEG
 KNSVPKRPVGPVYRLEMPNSPYASTDLRLSQRVHFNSPQEGYYPVSGVMSNSLSYSAVL
 DPVMVMVGEMVATRLFGTRVMDRFAYPDYTNLAGTSGPRMRRRIMQADKSLGRIFFFMFC
 CVVLCLLLF*

>G176 (41..1606)

AGAAGAAGAAGAAGAGTACCTCATACGTAAACCATTGATGGGCTCTTTTGATCGCCA
AAGAGCTGTTCCGAAATTCAAACAGCAACACCGTCACCGCTCCCTCTTTCTCCTTCGCC
TTACTTCACTATGCCTCCTGGCCTTACTCCCGCCGACTTTCTCGACTCTCCTCTTCTCTT
CACTTCCCTCAACATTTTGCCGTCTCCTACGACAGGCACATTTCCAGCGCAATCTCTGAA
CTATAACAATAACGGTTTGCTCATTGACAAAAATGAAATCAAATATGAAGACACAACCTCC
TCCCTTGTTCTTACCATCTATGGTAACCTCAGCCTTTACCTCAACTGGATTATTCAAATC
CGAAATCATGTGCGAGTAACAAAACCTCTGATGACGGCTACAATTGGCGCAAAATACGGGCA
GAAGCAAGTCAAAGGAAGCGAAAACCCGAGGAGTTACTTCAAATGCACGTATCCAAATTG
TCTCACAAGAAGAAAGTAGAGACGTCTCTTGTGAAGGGTCAGATGATTGAGATTGTCTA
TAAAGGAAGCCACAATCATCCCAAGCCCCAATCCACGAAGCGATCATCTTCCACCGCTAT
AGCAGCACATCAGAACAGCAGTAATGGAGACGGTAAAGACATTGGTGAAGATGAAACAGA
GGCCAAGAGATGGAAGAGAGAAGAGAAATGTGAAGGAGCCAAGAGTGGTGGTTCAGACAAC
AAGTGATATAGACATTCTTGACGATGGCTACAGATGGAGAAAGTATGGTCAGAAAGTCGT
CAAGGGTAATCCAAATCCAAGGAGCTATTACAAGTGCACATTTACAGGATGTTTTGTAAG
GAAACACGTTGAAAGAGCATTTCAGATCCCAAGTCAGTGATCACAACCTACGAAGGAAA
ACACAAACACCAAATCCCGACCCCAAGAGAGGTCCAGTTTAAAGATCTGCTGCAATGGC
TTCTCCTCTTCTCCCACTTCGACTACTCCTGATCAACTTCCCGCGGGCGATCCACAGTT
GCTGAGCTCTCTACGCGTCTCTTGTCCCGGTTCTAGCCACCGTCCGTACGCTTCTGC
AGATGCCAGACCTGGGCAGAGCTCGTTGACCGGTACGCTTTTCCCGGCCACCATCGCT
CTCGGAGGCAACGTACAGTAAGGAAGAACTTTTCTTATTTCCGAGCCAATTACATAAC
CTTAGTGGCAATCTTACTCGCCGCTCTCTGCTCAGCACCCCTTTTCTCTCTTCTCCT
CGCATCGCTGGCCGCTTCTTGCTTTTCTCTACTTTTCCGTCCGGCGGATCAGCCGTT
GGTCATTGGAGGACGACGTTCTCCGATCTTGAGACGCTAGGGATACTCTGCCTGTCCAC
TGTGGTGGTGGATGTTTATGACAGCGTTGGATCGCTCTTGATGTCCTACTAGCAGTTGG
GATCATGGGCGTGGCCATCCACGGAGCGTTTCTGCTCCCGAAGACCTGTTCTTGAAGA
ACAAGAAGCCATTGGATCTGGACTTTTTCGATTCTTCAACAACAATGCCTCTAATGCAGC
TGCCGCTGCCATAGCCACCTCAGCAATGTACGCGTTCGAGTCTGAGATTGTTGAAGAGA
CTACATTCTTACACCGCATTTCCAAAGTGTGATATTTATTCATATTGAATTGTT
>G176 Amino Acid Sequence (domain in AA coordinates: 117-173,234-290)
MGSFDRQRAVPKFKTATPSPLPLSPSPYFTMPPLTPADFLDSPLLFTSSNILPSPTTGT
FPAQSLNYNNGLLIDKNEIKYEDTTPPLFLPSMVTQPLPQLDLFKSEIMSSNKTSDDG
NWKYQKQVKGSSENPRSYFKCTYPNCLTKKKVETSLVKQMIEIVYKGSNHPKQSTK
RSSSTAIAAHQNSSNGDGKDIGEDETAKRWKREENVKEPRVVVQTTSDIDILDDGYRWR
KYGQKVVKGPNPNRSYKCTFTGCFVRKHVERAFQDPKSVITTYEGKHKHQIPTPRRGPV
LRSAMASPLPSTSTPDQLPGGDPQLLSLRLVLSRVLATVRHASADARPAELVDRSA
FSRPPSLSEATSRRVKNFSYFRANYITLVAILLAASLLTHPFALFLASLAASWFLYFF
RPADQLVIGGRTFSDLETGLILCLSTVVVFMFTSVGSLLMSTLAVGIMGVAIHGAFRAP
EDLFLEEQAIGSGLFAFFNNNASNAAAAAIATSAMSRVRV*
>G174 (194..1585)
CCCAATTGAGATTGTTTCGATTTTCGATCTACGAGATTCTTACAAGAACATAAGCAGCTTC
GGTTTTTGGGATTATCTTATTTGGTCGGATGATGATCTTCTCGATGTCTGTGCTAGGCT
TTGGGAATTAGATATATTTGGGGTTAAGCTCGAGTCTCTCCGGTTTTGAGTTTACTTGAG
TTTGTTAGTATTTATGGCTGAGGTGGGAAAAGTTCTGGCTAGTGATATGGAGTTAGACCA
TTCAAATGAGACTAAAGCAGTGGATGATGTTGTTGCCACTACTGATAAAGCGGAGGTCAT
ACCACTGGCTGTAAGTGAAGTGAACCGTTGTTGAAAGTTTGGAACTACTGACTGTAA
GGAGCTTGAAAACTTGTTCACATACGGTAGCTTCGCAGTCGGAAGTAGATGTTGCTTC
CCCGGTATCCGAGAAAGCACCGAAGGTTTCTGAAAGTAGCGGTGCATTATCTTTGCAGTC
TGGTTCGGAAGGGAATAGTCTTTTATTCTGTGAGAAGGTTATGGAAGACGGATACAACCTG
GCGGAAATATGGACAGAACTTGTAAGGAAATGAGTTTGTAAAGGAGCTATTACAGGTG
CACTCACCCCTAAGTGAAGCGGAAACAGTTGGAACGGTCTGCGGGTGGACAAGTCGT
GGATACCGTTTACTTTGGGGAAACATGATCACCCAAAGCCTCTTGCTGGTGCTGTTTCTAT
CAATCAGGATAAGCGAAGTGTGTTTACAGCTGTTAGTAAAGAGAAAAACATCTGGATC
CAGTGTTCAGACACTTCGTCAAACCGAACCACCAAGATCCATGGAGGATTACATGTTTC
AGTTATTCCACCAGCTGATGATGTGAAAACCTGATATTTACAATCAAGTAGGATAACGGG
GGACAACACTCACAAGGATTATAATAGTCCTACCGCAAAGCGAAGGAAGAAAGGAGGAA
CATTGAGCTGAGTCCAGTGGAGAGGTCAACCAATGATTCACGCATTGTGGTTTCACACTCA
GACTCTGTTTGATATTGTGAATGATGGGTACCGATGGCGTAAATATGGTCAGAAATCAGT

AAAAGGCAGCCCATATCCAAGGAGCTACTATAGATGTTCAAGCCCTGGATGCCCCGTCAA
 GAAACACGTAGAGAGGTCATCTCATGACACAAAGTTGCTTATAACAACCTACGAGGGAAA
 ACACGACCACGATATGCCTCCAGGAAGAGTTGTTACTCATAATAACATGCTGGACTCGGA
 AGTTGATGATAAAGAAGGAGATGCCAACAAGACTCCACAGAGCTCAACTCTTCAATCCAT
 TACAAAAGACCAGCATGTCGAAGATCACTTAAGAAAAGAAAACGAAGACTAATGGCTTTGA
 GAAAAGTCTTGATCAAGGTCCAGTTTGGATGAGAAGCTGAAGGAGGAAATAAAAGAGAG
 ATCAGATGCAAACAAGATCACGCAGCCAATCACGCCAAGCCGGAAGCAAAGTCAGATGA
 TAAAACCACTGTTTGTCAAGAGAAGGCAGTAGGAACCTGGAGAGCGAGGAACAAAACC
 CAAGACAGAGCCTGCCCAAAGCTAAGCATTCACTGTTGTACCGAGTGGTAATTTATATGG
 CTGTTTTAACATAGATTAGTACAGGCGATATGGTTATAGACTGTACAGTTGTTGTTTCAGG
 CGGGACCAGATTTAGATTAGTGTATTAATGGAATAGTATGCTTTAATACCTTTATGTAACC
 ACTTCCATTTGGTTCAAATAAGAGTTACAGGAAGAGAAGGTAACACAACAAGAGCCCTTC
 TTGTTGATGGAGCCTGTGTAATAGTTGTAGCATGGGGATGTATATGATTGATTCAACC
 TTATTAATGGTTATGAGACAAAACCTATC

>G174 Amino Acid Sequence (domain in AA coordinates: TBD)

MAEVGKVLASDMELDHSNETKAVDDVVAITDKAEVIPVAVTRTETVVESLESTDCKELEK
 LVPHTVASQSEVDVASPVSEKAPKVSESSGALSLSQSGSEGNPFIREKVMEDGYNWRKYG
 QKLVLKGFNFVRSYRCHPNCKAKKQLERSAGGQVVDTVYFGEHDPKPLAGAVPINQDK
 RSDVFTAVSKEKTSQSSVQTLRQTEPPKIHGGLHVSVIPPADVDKTDISQSSRITGDNTH
 KDYNSPTAKRRKKGNIELSPVERSTNDSRIVVHTQTLFDIVNDGYRWRKYGQKSVKGS
 YPRSYYRCSSPGCPVKHVERSSHDTKLLITTYEGKHDHMPGRVVTNNMLDSEVDDK
 EGDANKTPQSSTLQSIKTDQHVEDHLRKKTKTNGFEKSLDQGPVLDEKLKEIKERSDAN
 KDHAANHAKPEAKSDDKTTVCQEKAVGTLESEEQPKTEPAQS*

>G715 (1..705)

ATGGATACCAACAACAGCAACCACCTCCCTCCGCCGCCGGAATCCCTCCCTCCACCACCT
 GGAACCACCATCTCCGCCGCAGGAGGAGGAGCTTCTTACCACCACCTTCTCCAACAACAA
 CAACAACAGCTCCAACCTATTCTGGACCTACCAACGCCAAGAGATCGAACAAGTTAACGAT
 TTCAAAAACCATCAGCTTCCACTAGCTAGGATAAAAAAGATCATGAAAGCCGATGAAGAT
 GTTCGTATGATCTCCGCAGAAGCACCGATTCTCTTCGCGAAAAGCTTGTGAGCTTTTCATT
 CTCGAGCTCAGCATCAGATCTTGGCTTCACGCTGAGGAGAATAAACGTCGTACGCTTCAG
 AAAACGATATCGCTGCTGCGATTACTAGGACTGATATCTTCGATTTCTTGTGATATT
 GTTCCTAGAGATGAGATTAAGGACGAAGCCGCAGTCCTCGGTGTTGGAATGGTGGTGGCT
 CCTACCGCGAGCGCGCTGCCTTACTATTATCCGCCGATGGGACAACCAGCTGGTCTCTGGA
 GGGATGATGATTGGGAGAGACAGCTATGGATCCGAATGGTGTGTTATGTCCAGCCTCCGTCT
 CAGGCGTGGCAGAGTGTGTTGGCAGACTTCGACGGGGACGGGAGATGATGCTCTTATGGT
 AGTGGTGAAGTTCCGGTCAAGGGAATCTCGACGGCCAAGGGTAA

>G715 Amino Acid Sequence (domain in AA coordinates: 60-132)

MDTNNQPPPPSAAGIPPPPPGTTTISAAGGASVHLLQQQQQLQLFWTYQRQIEQVND
 FKNHQLPLARIKKIMKADEDVRMISAEAPILFAKACELFILELTIRSWLHABENKRTLQ
 KNDIAAAITRTDIFDLVDIVPRDEIKDEAAVLGGGMVVAPTASGVPIYYPPMPGQAPGPG
 GMMIGRPAMPDPNGVYVQPPSQAWQSVWQTSSTGTGDDVSYSGSGSGQGNLDGQG*

>G588 (196..1599)

ATCTGAAGTGAACCAAGCTCAGGTTTTGCTCTCTCTTTGATCATTCCTTTCTCAGCAATA
 TAAATTAGAGTTATATCCTTTATAAAGGATTTTGCTTTTTCACCAACAAACCCTAAATTC
 GGTGTCTCAGCAAGAATCACGTGATTCTCGTTCCTCTTCCTCAGCAAACCCATCATCTTC
 TATCTCATTTGAGAAATGGGTCAAAAGTTTTGGGAGAATCAAGAAGATCGAGCGATGGTT
 GAATCCACCATAGGCTCTGAAGCTTGGCACTTTTTCATCTCAACAGCTTCAGCTTCCAAC
 ACTGCCTTGTCCAAGCTTGTCTCACCACCAAGTGATTCCAATCTCCAACAAGGTTACGT
 CACGTTGTTGAAGGATCTGATTGGGATTATGCTCTTTTCTGGCTAGCGTCCAACGTTAAT
 AGCTCTGATGGTTGTCTCTTGATCTGGGGAGATGGTCATTGCCGTGTCAAAAAGGGTGCT
 TCAGGTGAGGATTACTCTCAGCAAGATGAGATCAAAAGACGTGTGCTTCGCAAGCTTCAC
 TTGTCTGTTGTTGGTTGAGATGAAGATCATCGTTTGGTGAAATCAGGAGCTCTTACTGAT
 CTCGACATGTTTTATCTGGCTTCTTTGTACTTTTCTTTAGGTGTGATACCAATAAGTAC
 GGTCTGCTGGAACCTATGTGTCTGGGAAGCCTCTTTGGGCTGCAGATTTGCCTAGCTGC
 TTGAGTTATTATAGGGTTAGGTCTTTCTTAGCTAGGTGAGCTGGTTTTCAGACTGTGTG
 TCTGTACCAGTGAATTCTGGAGTTGTGGAGCTTGGTTCTTTAAGACATATTCCAGAAGAT
 AAGAGTGTGATTGAGATGGTGAAATCAGTGTGTTGGTGGGTCTGACTTTGTTTCAGGCTAAA

GAAGCTCCTAAAATCTTTGGTCGACAGCTGAGTCTTGGTGGAGCAAAACCTCGGTCTATG
 AGTATTAATTTCTCCCGAAGACCGAGGATGACACGGGTTTCTCATTGGAATCGTATGAG
 GTGCAAGCGATCGGAGGCTCTAATCAAGTGTATGGTTATGAGCAAGGGAAAGATGAGACA
 TTGTATCTAACTGACGAGCAAAAGCCGAGGAAGAGAGGGGAGAAAACCAGCAAATGGAAGA
 GAAGAGGCTCTAAACCATGTGGAAGCGGAACGGCAGAGGAGGGAGAACTGAACAGAGA
 TTCTACGCTTTTGAGAGCGGTGGTGCCTAACATCTCCAAGATGGACAAGGCTTCGCTCCTT
 GCAGACGCAATCACTTACATCACGGATATGCAGAAGAAAATCAGGGTGTATGAAACAGAG
 AAGCAGATAATGAAGAGGAGGGAGAGTAATCAGATAACTCCAGCAGAGGTTGATTATCAA
 CAGAGGCATGATGATGCAGTTGTAAGGCTAAGCTGTCCGTTGGAACTCATCCAGTTTCA
 AAGGTGATACAAACGTTGAGGGAGAATGAAGTTATGCCTCATGATTCCAACGTGGCCATC
 ACAGAGGAGGGTGTGGTTACACATTCACCTCTCCGGCCTCAGGGTGGCTGCACCGCTGAG
 CAGTTGAAGGACAAGCTCCTTGCCTCTCTATCACAGTAACATATCACAGCAGTAACGTCTA
 TGTAATAAGTGTAACCGTGTGGAGGTTGTATCAATGTACTATTGCAAGCCAACCAAAAA
 AAACCTCCAGCTTAGTAGGATCGTGTAAATTTCTTATATGTAATGTTGAGATTGTCTTT
 TACATATAAAGATTGA

>G588 Amino Acid Sequence (domain in AA coordinates: 309-376)
 MGQKFWENQEDRAMVESTIGSEACDFFISTASASNTALSKLVSPPSDNLQQGLRHVVEG
 SDWDYALFWLASNVNSDGCVLWGDGHCVRVKGASGEDYSQQDEIKRRVLRKLHLSFVG
 SDEDHRLVKSGALTDLDMFYLASLYFSFRCDTNKYGPAGTYVSGKPLWAADLPSCLSYYR
 VRSFLARSAGFQTVLSVPVNSGVVELGSLRHIPEDKSVIEMVKS VFGGSDFVQAKEAPKI
 FGRQLSLGGAKPRSMSINFSPKTEDDTGFSLESYEVQAIIGSNQVGYEQKDETLYLTD
 EQKPRKRGRKPANGREEALNHVEAERQRREKLNRFYALRAVVPNISKMDKASLLADAIT
 YITDMQKKIRVYETEKQIMKRRESNQITPAEVDYQQRHDDAVVRLSCPLETHPVSKVIQT
 LRENEVMPHDSNVAITEEGVHTFTLRPQGGCTAEQLKDKLLASLSQ*

>G1758 (69..677)
 GTCCCTCCTCTTAGCTTCAACCGCCGGAATAAACAACCTTCTTGAAAAAAGAGA
 AACTAAAAATGAATATCCTTCAAACCTAACCTAGCTCCACAGATTTCACTGAATTTT
 TCAAGTTCGATGATTTTGACGATACTTTGGAGAAGATCATGGAAGAAATCGGCCGTGAGG
 ACCACTCGTCGTCACCGACTTTGAGTTGGAGTTCATCGGAAAAGTTAGTGGCTGCAGAAA
 TCACAAGCCCGCTTCAAACAAGCCTAGCTACCTCACCTATGAGCTTTGAAATAGGTGACA
 AAGATGAAATCAAAAAGAGGAAGAGACACAAAGAAGATCCGATTATTCACGTCTTCAAAA
 CGAAATCATCAATTGATGAAAAGGTTGCTTTAGATGATGGGTATAAATGGAGGAAATACG
 GAAAGAAGCCGATAACGGGTAGTCCATTTCCAAGGCATTATCACAGTGTTTCGAGCCAG
 ATTGCAACGTGAAGAAGAAGATCGAAAGAGATACGAACAATCCAGATTACATATTGACAA
 CATACGAAGGTAGACATAACCACCCAAGCCCTTCTGTAGTTTATTGTGATTACAGACGACT
 TTGATCTTAACCTCTCTCAACAATTGGTCCTTTTCAGACGGCAAATACGTATAGTTTCTCTC
 ATTCTGTCCATATTGATCGATCGTAGTTACAAGTTTGTGTATATAGATGTATATATATA
 TATACCAATTACCATCGTAATCACGTCTCACATGTAACGTACATATATCTTGTTC
 GGGGTTTCGTTTGTAAATGTATTGAATTGGTGGAGGTAGAATGGAAGTCATCTTGTATAGT
 TGTACTTGTATGTAAGTTTGATAGTCATTTTTTATAAAGTAACTAATTGTACAA

>G1758 Amino Acid Sequence (domain in AA coordinates: TBD)
 MNYPSPNPSPSTDFTEFFKFDDFDDTFEKIMEEIGREDHSSSPTLSWSSSEKLVAEITS
 PLQTSLATSPMSFEIGDKDEIKRKRHKEDPIIHVFKTKSSIDEKVALDDGYKWRKYGKK
 PITGSPFPRHYHKSSPDNCNVKKIERDTNPNPDYILTTYEGRHNHPSVSVYCDSDDFDL
 NSLNNWSFQTANTYSFSHSAPY*

>G2148 (66..737)
 GTCTCTAATATAAGCTTGAACGTTGCTATATATAAATGTAAAGGCGAACGCATAAGAAAA
 GAAAAATGGAGAATGAAGCTTTTGTAGATGGTGAATTGGAGTCTCTTTTGGGGATGTTCA
 ACTTTGATCAATTGTCATTAACGAATCGAGCTTTTGAATGCTCCAAATGAGACTGATG
 TTTTCTCTTCTGATGATTTCTCCATTTGGTACAATTCTGCAAAGTAACTATGCGGCCG
 TTCTTGATGGTTCCAACCACCAAACGAACCGAAATGCGACTCAAGACAAGATCTGTTGA
 AACCAAGGAAGAAGCAAAAGTTAAGCTCGGAAAGCAATTTGGTTACCGAGCCTAAGACTG
 CTTGGAGAGATGGTCAAAGCCTAAGCAGTTATAATAGTTTCAAGATGATGAAAAGGCTTTAG
 GTTAGTGTCTAATACATCAAAAAGCCTAAAACGCAAAGCGAAAGCCAACAGAGGGATAG
 CTTCGATCCTCAGAGCCTATACGCTAGGAAACGAAGAGAAAGGATAAACGATAGGCTAA
 AGACATTGCAGAGCCTAGTTCCCTAATGGGACAAAGGTGATATAAGCACAAATGCTGGAAG
 ATGCTGTCCATTACGTGAAGTTCCTGCAGCTTCAAATCAAGCTCTTGAGTTCAGAAGATC

TATGGATGTATGCACCTCTTGCTCACAATGGTCTGAATATGGGACTACATCACAATCTTT
TGTCTCGGCTTATTTAAGACAAAATCATTGGAATAACATAACTTACAGTACTTGTTTTTT
TTCTCGTTCTATATTCATGATTATGGTTATTTTTTGTGTTGAGTTGTTCAATTTTTCTGTC
TATTGCGTTCTATGAACCTGACACTCTTTTTGTAATTATTATATGCTAAAGACAATTTGG
ACTAACAGCATTTTAATAAAAAAAAAAAAA

>G2148 Amino Acid Sequence (conserved domain in AA coordinates:130-268)

MENEAFVDGELESLLGMFNFDQCSSNESSFCNAPNETDVFSDDFFPFGTILQSNYAAVL
DGSNHQTNRNVDNRQDLLKPRKKQLSSESNLVTEPKTAWRDGQSLSSYNSSDDEKALGL
VSNTSKSLKRKAKANRGIASDPQSLYARKRRERINDRLKTLQSLVPNGTKVDISTMLEDA
VHYVKFLQLQIKLLSSEDLWMYAPLAHNLNMGHLHNNLSRLI*

>G2379 (52..798)

CGCCGTCACCTCTCTCCCGGTGCCGCACATTAGCAACACTACTCCCGACGAATGGAGACG
ACGACGCCGCGAGTCAAAATCAAGTGTGTCCCACCGACCGCGTTGGGAAGAGAAGACTGG
TGGAGTGAGGAAGCGACGCGACGCTGGTAGAAGCCTGGGGCAATCGTTACGTCAAGCTG
AACCACGGAAATCTCCGGCAGAAATGACTGGAAAGACGTCGCCGACGCCGTTAACTCTAGA
CACGGTGATAACAGCCGTAAGAAGACCGACTTACAGTGTAAAGAACGGGTCGATACTTTG
AAGAAGAAGTACAAAACAGAGAAAGCTAAACTCTCGCCGTCGACTTGGCGTTTCTATAAC
CGCCTCGATGTTCTAATCGGTCCCGTTGTGAAGAAATCGGCTGGCGGAGTTGTCAAATCA
GCGCCTTTTAAAGATCATCTGAATCCAACCTGGATCGAATCTACTGGAAGCTCTCTTGAA
GATGATGATGAGGATGATGATGAGGTTGGTGATTGGGAATTCGTTGCTAGGAAGCATCCT
CGTGTGGAAGAGGTAGATCTGAGTGAAGGATCAACGTGTAGGGAAGTAGCTACGGCGATT
CTCAAGTTTGGAGAAGTTTACGAGAGAATTGAAGGAAGAAGCAACAGATGATGATTGAG
TTGGAGAAGCAGAGAAATGGAAGTGACAAAGGAGGTAGAGTTAAACGAATGAACATGTTG
ATGGAGATGCAGTTAGAGATTGAGAAATCAAAGCACCGGAAACCGCAAGTGCTTCAGGT
AAGAAGAACTCACATTAGG

>G2379 Amino Acid Sequence (domain in AA coordinates:19-110, 173-232)

METTTTPQSKSSVSHRPLGREDWWSEATATLVEAWGNRYVKLNHGNLRQNDWKDVADAV
NSRHGDNRSRKTDLQCKNRVDTLKKIKYTEKAKLSPSTWRFYNRLDVLIGPVVKKSAGGV
VKSAPFKNHLNPTGSNSTGSSLEDDDDDDDEVGDWEFVARKHPRVEEVDLSEGSTCRELA
TAILKFGEVYERIEGKKQMMIELEKQRMVTEKEVELKRMNMLMEMQLEIEKSKHRKRAS
ASGKKNSH*

>G1462 (63..1031)

CGTCGACCATTTCTTGCGATTGATCTTTCTCTAGATAATTTTTTTTGATCGATTTAGTTTCA
TTATGGAGGACGACGACGACGCTTATGATCTAATCAAACACGAAGTGTATCTCAGAAG
ACGAAGTAATAATCTCAGTTATCTGAAGGGTATGGTCGTTAACGGAGATTCTTGCCAG
ATCACTTCATCGAAGACGCAACGTTTACCAAGAATCCAGATAAGGTGTTCAATTCTG
AGAGACCTAGATTCTGTGATCGTTAAACCACGAACAGAGGCTTGTGGTAAACCGATGGAT
GTGATTCCGGGTTGCTGGAGGATCATTGGTCGTGATAAACTGATAAAGTCGGAGGAGACTG
GGAAGATTCTAGGGTTCAAGAAGATACTCAAGTTTGCCTAAAGAGGAAACCTATAGACT
ACAAAGAGAAGTTGGGTAATGGAAGATATAGGCTTACCAATAACTTGAAGTGAAGCAAG
ATCATGTGATTTGCAAAATTCGGTTTATGTTTGAAGCTGAAATTAGTTTCTTGCTAAGCA
AGCATTTCTACACTACATCAGAATCGGTTCTTGAATGAGCTGTTGCCATCTTATGGAT
ATTATTTATCCAATACACAAGAGGAGGATGAATTTTATCTGGACGCGATAATGACTTCGG
AAGGAAACGAGTGGCTAGCTACGTTACCAACAACGTGTACTGTCTGCATCCATTGGAGC
TTGTGGATCTTCAAGATCGGATGTTTAAATGATTACGGAACCTGCATCTTCGCTAACAGA
CTTGTGGTGAACTGATAAATGCGATGGTGCTTACTGGAAGATCCTGCACGGTGATAAGC
TGATCAAGTCAAATTTCCGAAAGGTCATTGGTTTCAAGAAGGTATTTGAGTTCTATGAAA
CGGTGAGACAAATATATCTTTGTGATGGAGAAGAAGTGACGGTAACCTTGGACTATACAAG
AGTATAGGCTTAGCAAAAACGTAAGCAGAATAAAGTGTGTGCGTTATCAAGTTGACTT
ATGATAGATAGGATACTTTACTTTGGTTTGTGATCATCTTAGTATCTTACGAATATTC
TAGATACACACATCTATAGGCGACCGCTCTAGACAGGCTCGTACCG

>G1462 Amino Acid Sequence (domain in AA coordinates: TBD)

MEDDDAAYDLIKHELLYSEDEVIIISRYLKMVVGNDSPDHFIEDANVFTKNPKVFNSE
RPRFVIVKPRTEACGKTGDCDSGCWRIIGRDKLIKSEETGKILGFKILKFCLKRKPIDY
KRSWVMEEYRLTNLNLWKQDHVICKIRFMFEAISFLLSKHFYTTSESVLENELLPSYGY
YLSNTQEEDEFYLDAIMTSEGNEWPSYVTNNVYCLHPLELVDLQDRMFNDYGTICIFANKT
CGETDKCDGGYWKILHGDKLIKSNFGKVIKVFYEFYETVRQIYLCDGEEVTVTWTIQE

YRLSKNVKQNKVLCVIKLTDR*

>G1211 (44..1120)

TGAAACCTAGATTTCTGCAACTGAATTCCTAATTCGAAAAAGAATGGAGGGTTCGTCGTC
 GACGATAGCAAGGAAGACATGGGAAC TAGAGAACAGCATTCTAACAGTAGACTCACCTGA
 TTCAACCTCCGACAACATCTTCTACTACGACGATACTTACAGACTAGGTTCCAGCAAGA
 GAAACCGTGGGAGAATGATCCTCACTACTTTAAACGAGTCAAGATCTCAGCGCTCGCTCT
 TCTTAAGATGGTGGTTCACGCTCGCTCTGGTGGTACAATTGAAATAATGGGTCTTATGCA
 AGGTAAGACCGATGGTGATACTATCATTTGTTATGGATGCTTTTGCTTTACCAAGTGAAGG
 TACTGAGACAAGGGTTAATGCTCAGGATGATGCTTATGAGTACATGGTTGAGTATTACACA
 GACCAACAAGCTCGCGGGGCTGGAGAATGTTGTTGGATGGTATCACTCTCACCCTGG
 ATATGGATGCTGGCTCTCCGGTATTGATGTTTCTACGACAGCGCTTAACCAACAGCATCA
 GGAGCCATTTTCTAGCTGTTGTTATTGATCCCAAGGACTGTTTCAGCTGGTAAGGTTGA
 GATTGGTGTCTTTCAGAACATACTCTAAAGGATATAAGCCTCCAGATGAACCTGTTTCTGA
 GTATCAAACCTATTCCTTTAAATAAGATTGAGGACTTTGGTGTTCAGTGCACAGTACTA
 TTCATTAGATGTCATTATTTCAAGTCATCTCTTGATTCTCACCTTCTGGATCTACTATG
 GAACAAGTACTGGGTGAACACTCTTTCTTCTCTCCACTGCTGGGTAAATGGAGACTATGT
 TGCTGGACAAATATCAGACTTAGCTGAGAAGCTTGAGCAAGCCGAGAGTCATCTGGTTCA
 GTCTCGCTTTGGAGGAGTTGTGCCATCATCCCTTCATAAGAAAAAGAAGATGAGTCTCA
 ACTCAATAAGATAACTCGGAGATAGCGCAAAGATAACTGTGGAACAGGTCCATGGACTAAT
 GTCGCAGGTATATAAAGATGAATTATTCAACTCAATGCGTCAGTCCAAACAACAAATCTCC
 CACTGACTCGTCGGATCCAGACCCTATGATTACATATTGAAGTTGCTCTTCTTTGGTTT
 CTANTTTTGGATTGACCCATCATTTGTTGCTCTTCTATTATTTTCTGTTGTGTAAAGAA
 TTATAATGNCGNCGCGAATTCGCGGCCGCTAAAAAANACAGGAAATTGAAAANAATTCTN
 NCCATTCCAACATCTTTATTTAATATTATCTCTCNATTATATAATATTCAAACATCCCT
 ANTANCTTCATTTGACCGTCCCCCTCCCTCCCGTGTTCNTTGGTGTCTGGCCCC

>G1211 Amino Acid Sequence (domain in AA coordinates: 123-179)

MEGSSSTIARKTWELENSILTVDSPDSTSDNIFYDDTSQTRFQQEKPWENDPHYFKRVK
 ISALALLKMOVHARSNGGTIEIMGLMQKTDGDTIIVMDAFALPVEGTETRVNAQDDAYEY
 MVEYSQTNKLAGRLNVVGYHSHPGYGCWLSGIDVSTQTLNQHQEPFLAVVIDPRTV
 SAGKVEIGAFRTYSKGYKPPDEPVSEYQTIPLNKIEDFGVHCKQYYSLDVITYFKSSLDSH
 LLDLLWNKYWVNTLSSSPLLGNGDYVAGQISDLAEKLEQAESHLVQSRFGGVVPSSLHKK
 KEDESQTLKITRDSAKITVEQVHGLMSQVIKDELFNSMRQSNKSPDSSDPDPMTY*

>G1048 (5..892)

GACCATGGCGGAGGAATTTGGAAGCATAGATTTACTCGGAGATGAAGATTTCTTCTTCGA
 TTTTCGATCCTTCAATCGTAATTGATTCCTTCCGGCGGAGGATTTTCTTCAGTCTTCACC
 GGATTCTAGGATCGGAGAAATCGAGAATCAATTGATGAACGATGAGAATCATCAAGAGGA
 GAGTTTTGTGGAATTTGATCAGCAATCGGTTTTCAGATTTCTAGCGGATCTACTCGTTGA
 TTATCCAACCTAGCGATTTCTGGCTCCGTTGATTGGCGGCTGATAAAGTTCTAACCGTCGA
 TTCTCCCGCCGCGCTGATGATTCCGGGAAGGAGAATTCCGATTGTTGTTGAGAAGAA
 GTCTAATGATTCCTGGTAGCGAGATTTCATGATGATGATGACGAAGAAGGAGACGATGATGC
 TGTGGCTAAAAACGAAGAAGGAGAGTAAGAAATAGAGATGCGGCGGTTAGATCGAGAGA
 GAGGAAGAAGGAATATGTACAAGATTTAGAGAAGAAGAGTAAGTATCTCGAAAGAGAATG
 CTTGAGACTAGGACGTATGCTTGAGTGCTTCGTTGCTGAAAACCAGTCTCTACGTTACTG
 TTTGCAAAAGGGTAATGGCAATAATACTACCATGATGTGCAAGCAGGAGTCTGCTGTGCT
 CTTGTTGGAATCCCTGCTGTTGGGTTCCCTGCTTTGGCTTCTGGGAGTAACTTCATTG
 CCTATTCCCTTATATGTCCACACAAAGTTGTCCTCTACGTCCAGAACAGAAAAGCT
 GGTCTAAACGGGCTCGGGAGTAGTAGCAAACCGTCTTATACCGGCGTTAGTCGGAGATG
 TAAGGGTTTCGAGGCCTAGGATGAAATACCAAATCTTAACCTTGCAGCGTGACAACGCCT
 TTTTAACTGCTTCTTTTGCATTTTGGATTGTAGATGAGTGTCTTTTAGTTTTCTCTC
 TCTTGTGTTTGTATTTCGCTGTTGAAAGTTTCTGTCTAATATCGATAAGTTAACAGTGAA
 AAAAAAAAAAAAAA

>G1048 Amino Acid Sequence (domain in AA coordinates 138-190)

MAEEFGSIDLLGDEDFFFDFDPSIVIDSLPAEDFLQSSPDWIGEIEENQLMNDENHQEES
 FVELDQQSVSDFIADLLVDYPTSDSGSVDLAADKVLTVDSPAAADDSGKENS DLVVEKKS
 NDSGSEIHDDDDDEEGDDAVAKRRRRVRNRDAVRSRERKKEYVQDLEKSKYLERECL
 RLGRMLECFVAENQSLRYCLQKNGNNTTMMSKQESAVLLLESLLLSLLWLLGVNFICL
 FPYMSHTKCCLLRPEPEKLVNLGLSSSKPSYTGVSRRCKGSRPRMKYQILTLAA*

>G986 (31..846)

CATTAAATTGGCTCCTGTGAACCTAAATTTATGGACTATGATCCCAACACCAATCCGTTT
 GACCTTCATTTCTCCGGTAAACTTCCGAAAAGAGAAGTCTCGGCTTCAGCTTCTAAAGTT
 GTAGAGAAGAAATGGTTAGTGAAAGATGAGAAGAGAAATATGCTACAAGATGAAATAAAC
 CGGGTTAATTCGGAGAACAGAAGCTAACCGAAATGTTAGCAAGAGTCTGTGAGAAGTAC
 TATGCTCTTAATAATCTTATGGAGGAGTTGCAGAGTCGAAAGAGTCTGAAAGTGTTAAC
 TTTTCAGAACAAACAGCTAACGGGGAAACGAAAACAAGAAGTCTGATGAGTTTGTTAGCTCC
 CCAATTGGACTCAGTCTCGGACCAATCGAGAACATCACCAACGATAAAGCGACGGTTTCA
 ACCGCTTACTTTGCTGCTGAGAAGTCTGACACAAGCTTGACTGTGAAAGATGGATATCAA
 TGGAGGAAATACGGGCAAAAGATTACGAGAGATAATCCATCTCCTAGAGCTTACTTCAGA
 TGCTCGTTTTACCGTCTTGCTAGTCAAGAAGAAGGTGCAACGAAGTGCAGAAGATCCA
 TCTTTCTTGGTAGCCACTTACGAAGGGACACATAACCACACCGGACCACATGCAAGTGTG
 TCCAGGACAGTGAAACTTGATCTAGTTCAAGGTGGGCTTGAACAGTTGAGGAAAAGAAA
 GAGAGAGGGACGATTCAAGAGGTTTTGGTGCAACAAATGGCTTCTTCGTTGACCAAAGAT
 CCTAAGTTCACTGCAGCTCTTGCGACTGCTATTTCCGGGAGATTGATAGAGCATTCAAGA
 ACATGAAAGTTCTCTAGAACATGTATATTTCTGTTTGTCTATTTTGTGCTCATTCCCT
 AGTAAAAAGGTAAAGATTTGTTTGATCTTGATTAGGAGGCATAGATGTCAATTTTAATGT
 GTGTGTATATAATTACATCAATCTAAGTATCCAAAAAGGGTCACCCCCATTTTATCTTA
 TG

>G986 Amino Acid Sequence {domain in AA coordinates: 146-203}

MDYDPNTNPFDLHFSGLPKREVSAASAKVVEKKWLVKDEKRNMLQDEINRVNSENKKLT
 EMLARVCEKYYALNNLMEELQSRKSPESVNFQNKQLTGKQKQELDEFVSSPIGLSLGPIE
 NITNDKATVSTAYFAAEKSDTSLTVKDGQWRKYQKQITRDNPSPRAYFRCSFSPSCLVK
 KKVQRSAEDPSFLVATYEGTHNHTGPHASVSRVTKLDLVQGGLEPVEKKERGTTIQEVLV
 QQMASSLTKDPKFTAAALATAISGRLEHSRT*

>G789 (259..1593)

GGCAAGAAGAACCTTAGCCTCTCTTTCTTCTTCTCTCTCTCTCTCTGTGGTACTGTT
 CTGTTTCAACTTTACTCCCTCAGTTTCAGAACAAATCCCTATCTAGAAGAGAGATAAAAC
 CGAGAAGGTTTTGGAGATAGAATCTTTGTTCTTCTTTTGTCCCTCCTTGCTCGATTTTT
 GTTACGTGTGAAGCAATAAAAAAACTGATATAGCTAAATCTTCCATCCATTACAGAGGC
 TTCTAAATCTGATCTGACATGGAACAAGTGTGCTGATTGGAATTTTGAAGATAATTTT
 CACATGTCCACTAATAAAAGATCAATCAGACCAGAAGATGAATTAGTGGAGCTATTGTGG
 AGAGATGGTCAAGTGGTTTTACAAAGCCAAGCTCGTAGAGAACCGTCAGTCCAAGTCCAA
 ACCACAAAACAAGAAACCTTAAGAAACCCCAACAATATTTTCTTGACAACCAAGAAACA
 GTACAAAAGCCTAACTACGCTGCTCTAGATGATCAAGAAACCGTCTCCTGGATACAATAC
 CCTCCGGATGACGTCATCGACCTTTTGAATCCGAGTCTCCTCTCATTCTTCTCTCTCG
 ATCGATCACCTCGGAGGTCTTGAGAAGCCACGAACGATCGAAGAGACAGTTAAGCATGAG
 GCTCAAGCCATGGCTCCTCCTAAGTTTAGATCCTCGGTTATAACAGTCGGACCGAGTCAT
 TCGCGCAGCAACCAAGTCAACAAATATTCATCAGGCCACTACACTTCCGGTTTCTATGAGT
 GATAGAAGCAAGAACGTGGAAGAAAGACTTGACACTTCGTCAGGTGGCTCCTCCGGTTGC
 AGCTATGGAAGGAACAACAAGAAACCGTTAGTGAACAAGTGTAACCATTGACCGTAAA
 AGAAAACATGTTATGGATGCTGATCAAGAATCTGTGCTCAATCAGATATAGGTTTGACC
 TCAACCGATGATCAAAACCATGGGTAAACAATCGAGCCAACGGTCAGGATCTACTCGAAGA
 AGCCGTGCAGCTGAAGTTTATAATCTCTCAGAAAGGAGGAGGAGAGATCGGATCAATGAA
 AGAATGAAAGCTCTTCAAGAACTCATACCTCACTGCAGCAGAACAGATAAAGCTTCGATA
 TTGGATGAAGCAATTGATTACTTAAATCACTTCAAATGCAACTCCAAGTGATGTGGATG
 GGAAGTGAATGGCGGCGGCGGAGCAGCAGCAAGTCCGATGATGTTTCCCGGGGTA
 CAATCATCTCCATACATTAATCAGATGGCTATGCAAAGTCAGATGCAATTGTCTCAATTC
 CCGGTTATGAACCGGTCCGCTCCGCAGAACCATCCCGGTTTAGTATGTCAAAACCCGGTA
 CAGTTGCAGCTCAAGCACAGAACCAATCTTATCGGAGCAGCTCGCTAGGTACATGGGC
 GGGATTCCCGAGATGCCCGCGGCGGAAATCAGATGCAGACCGTGCAACAACAACAGCG
 GACATGTTGGGATTTGGATCTCCGGCGGGACCGCAAAGTCAACTGTCCGCACCGGCGACC
 ACCGACAGTCTTCATATGGGTAAATAGGCTGACTTGGCATATAGTTTTCTCCGAAATT
 ATTCTTCTTACAGTTGGTGATTGTTATTTATTTTGGTCGCCTAAGCAAGCATAAAAGCT
 AAGTCAAATGTATTATAGAGATCTAATAAGTTAGTCTCATACTTATAACTTATTTTTAAA
 CAGTTGAATTATAGTATCAATCAAGTGTGGGAACCTAAGATCATACATGTGTCAATAC
 TTTTATATTTGTTCTCAAGGTTTCATCAGAAAAACAAATAAAAAGGATAGACTAGGCCTG

CATTGACATTATCATGGGCTTTTTGGGTCTATGAATATGAACATTAACCCC
>G789 Amino Acid Sequence (domain in AA coordinates: 253-313)
MEQVFADWNFEDNFHMSTNKRISIRPEDELVELLWRDQGVVLSQARREPSVQVQTHKQET
LRKPNNIFLDNQETVQKPNYAALDDQETVSWIQYPPDDVIDPFSEFSSHFSSIDHLGG
PEKPRTIETVKHEAQAMAPPKFRSSVITVGP SHCGSNQSTNIHQATTLPVMSDRSKNV
EERLDTSSGGSSGCSYGRNNKETVSGTSVTIDRKRKHVMDADQESVSQSDIGLTSTDDQT
MGNKSSQSRSGSTRRSRAAEVHNLSERRRRDRINERMKALQELIPHCSRDKASILDEAID
YLKSLQMLQVMWMSGMAAAAAAASPMMPGVQSSPYINQMAMQSQMQLSQFPVMNRS
APQNHPLVLCQNPVQLQLQAQNLSEQLARYMGGIPQMPPAGNQMQTVQQPADMLGFG
SPAGPQSQLSAPATTDLSLHMGKIG*

>G2085 (1..930)

ATGTTTGGTGCCTTCGATTATCCCAAATAACCAGATTGGTACCGCCTCTGCTTCCGCT
GGTGAAGACCATGTCTCTGCCCTCCGCTACGCTCGGTACATTCTTACGACGATATGGAA
GAAATCCCTCATCTGACTCTATCTATGGTGTCTGCCCTCCGATTGATTCCCGATGGCTCT
CAATGGTTGCTCACCGATCCGATGGCTCTGAATTACTTGTCTCGGCCACCGGAAGGG
GCGAATCAGCTTACGATCTCGTTCCGTGGACAAGTTTACGTTTTTGATGCCGTTGGTGTCT
GACAAGGTGGATGCTGTGTTGTGCTGTTGGGTGGTTCTACTGAGCTTGCTCCTGGTCCG
CAGGTGATGGAAGTACTCAACAGCAGAATCATATGCCTGTTGTAGAATATCAGAGCCGC
TGTAGCCTTCCGCAACGGGCACAATCCTTGGATAGGTTTCGGAAGAAGAGGAATGCTAGA
TGTTTCGAGAAGAAAGTAAGATACGGTGTTCGCCAAGAAGTTGCCTTAAGAATGGCACGT
AATAAAGGTCAATTCACCTCTTCAAAGATGACAGATGGGGCTTATACTCTGGCACAGAT
CAAGATTCTGCCCAAGATGATGCCCATCCAGAAATATCGTGTACTCATTGCGGCATTAGT
TCCAAATGTACCAATGATGCGACGTGGCCCTTCCGGCCCCAGGACTCTTGCAATGCC
TGTGGACTTTTTTGGGCTAACAGGGGTACATTGAGGGATCTCTCAAAGAAAACAGAAGAG
AATCAGTTGGCTTTAATGAAACCGGATGATGGTGGGAGTGTGCTGATGCTGCTAACAAC
TTAAACACTGAAGCTGCAAGTGTGAAGAACACACTTCCATGGTTTCTCTTGCCAATGGG
GATAATTCTAATCTGTTAGGTGATCACTAA

>G2085 Amino Acid Sequence (domain in AA coordinates: TBD)

MFGRHSIIPNNQIGTASASAGEDHVSASATSGHIPYDDMBEIPHPDSIYGAASDLIPDGS
QLVAHRSDGSELLVSRPPEGANQLTISFRGQVYVFDVAGADKVDVLSLLGGSTELAPGP
QVMELAQQQNHMPVVEYQSRCSLPQRAQSLDRFRKKRNARCFEKKVRYGVRQEVALLMAR
NKGQFTSSKMTDGAYNSTGTDQDSAQDDAHPEISCTHCGISSKCTPMMRRGPGSPRTLCA
CGLFWANRGTLRLDLSKKTEENQLALMKPDDGGSVADAANNLNTAASVEHTSMVSLANG
DNSNLLGDH*

>G1783 (1..603)

ATGGCCGCGTTTCCGCGAGTGGACAAGGGTCGATGACAAACGTTTTGAGTTAGCTCTGCTT
CAAATCCCGGAGGGTTCGCCGAATTTATAGAGAAATATCGCCTATTATCTCCAGAAACCG
GTGAAGGAGGTGGAGTACTACTGCGCGTTGGTCCATGATATTGAGCGGATCGAATCG
GGTAAGTATGTTTTGCCCAAATACCCGGAAGACGATTACGTGAACTGACGGAAGCAGGT
GAGTCTAAGGGCAATGGGAAAAGACGGGAATTCCTTGGTCAGAAGAGGAACAGAGGTTG
TTTCTGGAAGGACTAAATAAGTTTGGGAAAGGAGACTGGAAGAACATATCGAGGTATTGT
GTGAAGTCAAGGACCTCGACGCAAGTGGCAAGCCATGCTCAGAAGTATTTTGCAAGGCAA
AAGCAGGAGAGTACGAATACTAAACGCCCGAGTATTCATGACATGACTCTGGGAGTTGCG
GTCAATGTCCCTGGATCCAATTTGGAGTCTACTGGCCAGCAACCACATTTTGGTGATCAA
ATTCCTTCGAATCAATATTATCCCTCCAGGAAACTTTTCGGGGTTTTGATCAGCGATGG
TGA

>G1783 Amino Acid Sequence (domain in AA coordinates: 81..129)

MAAFPQWTRVDDKRFELALLQIPEGSPNFIENIAYYLQKPVKEVEYYYCALVHDIERIES
GKYVLPKYPEDDYVKLTEAGESKGNKKTGIPWSEBEQRLFLEGLNKFKGDKWKNISRYC
VKSRTSTQVASHAQKYFARQKQESTNTRPSIHDMTLGVAVNVPGSNLESTGQQPHFGDQ
IPSNQYYPSQENFRGFDQRW*

>G2072 (155..793)

TCGACCCACGCGTCCGCCACGCGTCCGGATCTTTTCACAGAAGACCAACCAGCTTGGCT
CGATGAGCTCCTAAGTGAGCCAGCATCACCTAAGATTAACAAAGGTCATAGACGTTTCAGC
TAGTGACACAGCTGCTTACTTGAACCTCAGCTTTAATGCCTTCGAAGGAAAATCATGTTGC
TGGTTCGTCTTGGCAGTTCAGAACTATGATTTGTGGCAGTCCAACCTTATGAACAACA
CAATAAATTAGGATGGGATTTCTCTACAGCAAATGGAACATAATATCCAAGAAATATGTC

ATGCGGAGCTTTAAATATGTCGTCGAAACCCATTGAGAAACATGTAAGCAAAATGAAAGA
AGGAACCTTACAAAACCAGATGGTCTTAGATCAAAGACTGACTCAAAACGTATCAAACA
TCAAAATGCTCATCGAGCGCGTTTGAGAAGGCTTGAGTACATATCAGACCTTGAAAGGAC
CATCCAAGTGCTACAAGTTGAAGGATGTGAAATGTCATCTGCCATTCACTACTTGGATCA
GCAGTTACTCATGCTTAGCATGGAAAAATAGAGCTTTAAACAACGTATGGATAGTTTAGC
AGAAATCCAAAAGCTTAAACATGTGGAGCAGCAATTGCTTGAGAGAGAGATAGGAAACCT
ACAGTTTCGACGACACCAACAACACACAGCAAAACCAAAAACAAGTCCAAGCAATACA
AAATCGATACACCAATATCAACCACCTGTTACACAAGAACCCGATGCCCAATTTGCAGC
CTTGGCAATATGATTTAGGAAATATGGATACATTGTTTCAGATTAAGCTGAGCTCCTCTTG
CTCTACCTTAATGTCCATACAACATAGGTGAACCTTGATGTTTGTAGCCTTGAATGAAAC
CTAAAAAGCATCGTTATGTAAATCAAATGTGGTTGCCCATATCCTCCTCTATTGCATT
TCTCTCTATTATGGCATGGTAGAGAACTCTTGTCAAGAACTTCATGTTATGTAATAA
CTTGTAATCCTTCTTATTTTCATCTATTATATATATGAATAAGTAATTTTTTTGCCAAAAA
AAAAAAAAAAAAAAAAAAAA

>G2072 Amino Acid Sequence (conserved domain in AA coordinates:90-149)

MPSKENHVAGSSWQFQNYDLWQNSYEQHNKLGWDFSTANGTNIQRNMSCGALNMSSKPI
EKHVS KMKEGTSTKPDGPRSKTDSKRIKHQNAHRARLRRLLEYISDLERTIQVLQVEGCEM
SSAIHYLDQQLMLSMENRALKQRMDSLAETQKLKHVEQQLLEREIGNLQFRRHQQQPQQ
NQKQVQAIQNRYTKYQPPVTQEPDAQFAALAI*

>G931 (85..1071)

GGAGGTTCTTTGACAGACACATGTATCATCAATCTTCTCTGTTGAAGCAGAGAGAGAGAG
AGCTAATTGTTGCCTCTGAGTCACATGGATAAGAAAGTTTCATTTACTAGCTCTGTGGCA
CATTCAATCCACCATACTTTAGTACTTCCATCTCATGGGACTTCCAACCAATCCAAT
GGTGTGACTGAATCACTGAGTTTGAAGGTGGTAGATGCAAGACCAGAACGTCTTATAAAC
ACAAAGAATATCAGTTTCCAGGACCAGGATTCATCTTCAACTCTGTCTCTGTCTCAATCT
TCTAACGATGTTACAAGTAGTGGAGATGATAACCCCTCAAGACAAATCTCATTTTATAGCA
CATTGAGATGTTGTAAAGGATTTGAAGAACTCAAAGGAAGCGATTTGCAATTAAATCA
GGCTCCTCCACGGCAGGAATCGCTGATATCACTCTTCTCTCTCCAAGGCTAACTTCTCA
TTTCACTATGCCGATCCACATTTTGGTGGTTTAAATGCCCTGCGGCTTACCTACCACAGGCA
ACAATATGGAATCCCCAAATGACTCGAGTTCCGCTACCATTCCGATCTCATAGAGAATGAG
CCTGTCTTTGTCAATGCAAAGCAATTCCATGCAATTATGAGGAGGAGGCAACAGCGTGCT
AAGCTAGAGGCGCAAAACAACATAATCAAAGCCCGTAAGCCGATCTTTCATGAATCTCGA
CATGTTACAGCTCTTAAACGACCTAGAGGATCTGGTGGAGATTCTTAAACACCAAAAAG
CTTCAAGAATCTACAGATCCAAAACAAGACATGCCAATCCAACAGCAACACGCAACGGGA
AACATGTCAAGATTTGTGCTTTATCAGTTGCAGAACAGCAATGACTGTGATTGTTCAACC
ACTTCTCGCTCTGACATCACATCTGCTTCTGACAGCGTTAATCTCTTTGGACACTCTGAA
TTTCTGATATCAGATTGCCCATCTCAGACAAACCAACATGTATGTTTCATGGTCAATCA
AATGACATGCATGGAGGTAGGAACACACACCATTTCTCTGTCCATATCTGAGCCGGTGGA
ATCTGGTAATGTGTACGTTTCTACAAAAAAGGGAAGTCATCCTTGGCTGCTACTTTCGCT
TATTAGCTAGTCTTATTTTACACGCTTTGTCCAGATATC

>G931 Amino Acid Sequence (domain in AA coordinates: TBD)

MDKKVSFTSSVAHSTPPYLSISWGLPTKSNVTESLSLKVVDARPERLINTKNISFQD
QDSSSTLSSAQSSNDVTSSGDDNPSRQISFLAHSVDCKGFEEQQRKFAIKSGSSTAGIA
DIHSSPSKANFSFHYADPHFGGLMPAAYLPQATIWNPMQTRVPLPFDLIENEPVFNNAKQ
FHAIMRRRQRAKLEAQNKLIKARKPYLHESRHVHALKPRGSGGRFLNKKLQESTDPK
QDMPQQQHAATGNMSRFVLYQLQNSNDCDCSTTSRSDITSASDSVNLFGHSEFLISDCPS
QTNPTMYVHGQSNMDMHGGRNTHHFSVHI*

>G278 (93..1874)

TCGATCTTTAACCAATCCAGTTGATAAGGTCTCTTCGTTGATTAGCAGAGATCTCTTTA
ATTTGTGAATTTCAATTCATCGGAACCTGTTGATGGACACCATTGATGGATTCGCCG
ATTCTTATGAAATCAGCAGCACTAGTTTCGTCGCTACCGATAACACCGACTCCTCTATTG
TTTATCTGGCCGCCGAACAAGTACTCACCGGACCTGATGTATCTGCTCTGCAATTGCTCT
CCAACAGCTTCGAATCCGTCTTTGACTCGCCGGATGATTTCTACAGCGACGCTAAGCTTG
TTCTCTCCGACGGCCGGGAAGTTTCTTCCACCGGTGCGTTTGTGTCAGCGAGAAGCTCTT
TCTTCAAGAGCGCTTTAGCCGCCGCTAAGAAGGAGAAAGACTCCAACAACACCGCCGCCG
TGAAGCTCGAGCTTAAGGAGATTGCCAAGGATTACGAAGTCGGTTTCGATTCCGGTTGTGA
CTGTTTTGGCTTATGTTTACAGCAGCAGAGTGAGACCGCCGCTAAAGGAGTTTCTGAAT

GCGCAGACGAGAATTGCTGCCACGTGGCTTGCCGGCCGGCGGTGGATTTCATGTTGGAGG
 TTCTCTATTTGGCTTTTCATCTTCAAGATCCCTGAATTAATTACTCTCTATCAGAGGCACT
 TATTGGACGTTGTAGACAAAGTTGTTATAGAGGACACATTGGTTATACTCAAGCTTGCTA
 ATATATGTGGTAAAGCTTGTATGAAGCTATTGGATAGATGTAAAGAGATTATTGTCAAGT
 CTAATGTAGATATGGTTAGTCTTGAAGAGTCATTGCCGGAAGAGCTTGTAAAGAGATAA
 TTGATAGACGTAAAGAGCTTGGTTTGGAGGTACCTAAAGTAAAGAAACATGTCTCGAATG
 TACATAAGGCACTTGACTCGGATGATATTGAGTTAGTCAAGTTGCTTTTGAAGAGGATC
 ACACCAATCTAGATGATGCGTGTGCTCTTCATTTGCTGTTGCATATTGCAATGTGAAGA
 CCGCAACAGATCTTTTAAACCTTGATCTTGCCGATGTCAACCATAGGAATCCGAGGGGAT
 ATACGGTGTCTCATGTTGCTGCGATGCGGAAGGAGCCACAATTGATACTATCTCTATTGG
 AAAAAGGTGCAAGTGCATCAGAACCACTTTGGAAGGTAGAACCAGCACTCATGATCGCAA
 AACAAGCCACTATGGCGGTTGAATGTAATAATATCCCGGAGCAATGCAAGCATTTCTCTCA
 AAGGCCGACTATGTGTAGAAATACTAGAGCAAGAAGACAAACGAGAACAATTCCTAGAG
 ATGTTCTCCCTCTTTTGCAGTGGCGGCCGATGAATTGAAGATGACGCTGCTCGATCTTG
 AAAATAGAGTTGCACTTGCTCAACGTCTTTTCCACGGAAGCACAAGCTGCAATGGAGA
 TCGCCGAAATGAAGGGAACATGTGAGTTCATAGTGACTAGCCTCGAGCCTGACCGTCTCA
 CTGGTACGAAGAGAACATCACCGGGTGTAAAGATAGCACCTTTCAGAATCCTAGAAGAGC
 ATCAAAGTAGACTAAAAGCGCTTTCTAAAACCGTGGAACCTCGGGAACGATTCTTCCCGC
 GCTGTTCCGGCAGTGTCTGACCAAGATTATGAAGTGTGAGGACTTGACTCAACTGGCTTGCG
 GAGAAGACGACACTGCTGAGAAACGACTACAAAAGAAGCAAAGGTACATGGAAATACAAG
 AGACACTAAAGAAGGCCCTTAGTGAGGACAATTTGGAATTAGGAAATTCGTCCTTGACAG
 ATTCGACTTCTTCCACATCGAAATCAACCGGTGGAAGAGGTCTAACCGTAAACTCTCTC
 ATCGTCGTCGGTGAGACTTTGCCTCTTAGTGTAATTTTGTCTGTACCATATAATTCTGT
 TTTTCATGATGACTGTAACTGTTTATGTCTATCGTTGGCGTCATATAGTTTCGCTCTTCGT
 TTTGCATCCTGTGTATTATTGCTGCAGGTGTGCTTCAAACAAATGTTGTAAACAATTTGAA
 CCAATGGTATACAGATTTGTAATATATATTTATGTACATCAACAATAAAAAAAAAAAAAA
 AAAA

>G278 Amino Acid Sequence (domain in AA coordinates: 2-593)
 MDTTIDGFADSYEISSTSFVATDNTDSSIVYLAEEQVLTGPDVLSALQLLSNSFBSVFDSP
 DDFYSDAKLVLSDGREVSFHRCVLSARSSFFKSALAAAKKEKDSNNTAAVKLELKEIAKD
 YEVGFDSVVTVLAYVYSSVRPPPKGVSECADENCCHVACRPVDFMLEVLYLAFIFKIP
 ELITLYQRHLLDVVDKVVIEDTLVILKLANICGKACMKLLDRCKEIIVKSNDVMVSLEKS
 LPEELVKEIIDRRKELGLEVPKVKHVS NVHKALDSDDIELVKLLKEDHTNLDACALH
 FAVAYCNVKTATDLLKLDLADVNHRNPRGYTVLHVAAAMRKEPQLILSLLEKGASASEATL
 EGR TALMIAKQATMAVECNNIPEQCKHSLKGRLCVEILEQEDKREQIPRDVPPSFAVAAD
 ELKMTLLDLENRVALAQR LFPTEAQAAMEIAEMKGTCEFIVTSLEPDRLTGKRTSPGVK
 IAPFRILEEHQSRLKALSKTVELGKRFFPRCSAVLDQIMNCEDLTQLACGEDDTAEKRLQ
 KKQRYMEIQETLKKAFSEDNLELGNSSLTDSTSSSTSKSTGGKRSNRKLSHRRR*

>G2421 (1..630)
 ATGGAGGGTTTCGTCCAAAGGGTTGAGGAAAGGTGCATGGACTGCTGAAGAAGATAGTCTC
 TTGAGGCACTGTATTGGTAAGTATGGAGAAGGCAAATGGCATCAAGTTCTTTAAGAGCT
 GGGCTAAATCGGTGCAGGAAAAGTTGTAGACTAAGATGGTTAAACTATTTGAAGCCAAGT
 ATCAAGAGAGGAAAATTTAGTTCTGATGAAGTTGATCTTCTTCTTCGTCTTCATAAGCTT
 CTAGGAAATAGGTGGTCCCTTGATTGCTGGTCGATTACCTGGTCCGACCGCTAATGATGTC
 AAGAACTACTGGAACACCCATCTGAGTAAGAAGCATGAACCGTGTGTAAAACATAAGATA
 AAAAGGATAAATATTATAACCCCTCCTAATACACCGGCCCAAAAAGTTTGTGAAAATAGT
 ATCATATGTAACAAAGATGATGAGAAAGATGATTTTGTGGATAATTTTATGGTTGGAGAT
 AATATATGGTTGGTGGAGCTTTGTCTAGACGAGGGCCAAGAGGTAGATGTGCTGGTTACAGAA
 GCGGCGGCAACAGAAAAGGAGGGCACTTTGGCGTTTGACGTTGAGCAACTTTGGAATTTG
 TTCGATGGAGAGACTGTGATCTTTGATTAGTGTTTATAAACGTTTGTGTTCTCTTGTGTTG
 TGAGGTTTCTCTATTTTAATTTAGTATCTATTTTCTAAATTAATAATATCTTATAGTATT
 TTAGGCAAACCTTATGTTTCCGTTTCTGTGCGGCCGCTCTAG

>G2421 Amino Acid Sequence (domain in AA coordinates: 9-110)
 MEGSSKGLRKGAWTAEDSLLRQCIGKYEGEKWHQVPLRAGLNRCRKSRLRLWNLNLYLKPS
 IKRKGFSSDEVDLLRLHLKLLGNRWSLIAGRLPGRTANDVKNYWNTHLSKKHEPCCKTKI
 KRINII TFPNTPAQVKCENSITCNKDDEKDDFVDNFMVGDNIWLERLLDEGQEVLDVLVTE
 AAATEKEGTLAFDVEQLWNLFDGETVIFD*

>G2032 (53..1789)

TCCCTCCCAGAGTAAGAACTTCCATACTTTGCTCTAGATTTCTTGAGAAAAGATGCAGCC
 GATCTTCCATGCGATCCTTAAAAATGACCTTCCAGCTTTTITAGAGTTGGTAGAAGATAG
 TGAATCGTCTCTGGAGGAGAGAAAACGAGGAAGAACACTTGAACAACACGGTTTTGCACAT
 GGCTGCAAAGTTTGGTCACCGAGAACTCGTCTCCAAGATTATTGAGCTCCGACCTTCCCT
 CGTGTCTTCCCGCAACGCATACAGAAACACACCTTTGCATCTTGCTGCTATCCTTGGAGA
 TGTAACATAGTTATGCAGATGTTAGAGACTGGATTGGAAGTGTGTTCTGCACGCAATAT
 CAACAACCACACACCACTCCACTTGGCTTGCCGTAGCAATTCCATAGAGGCTGCCAGACT
 CATCGCGGAAAAGACACAATCAATTGSCCTCGGTGAACTCATTCTCGCCATATCAAGTGG
 ATCCACTAGTATCGTAGGGACTATACGGAGAGATTCCCAGACCTAGCTAGGGAAGAAGC
 TTGGGTGGTTGAAGACGGCTCACAATCAACGCTACTGCATCATGCGTGTGATAAGGGAGA
 CTTTGAAGTGAACATATATTGTTAGGGCTCGATCAAGGATTAGAAGAAGCACTTAACCC
 CAATGGTTTATCACCTCTGCATCTTGCGGTCTCAGAGGCTCGGTTGTGATCCTGGAGGA
 GTTCTTGGACAAGGTTCCATTGTCTTTCAGCTCAATCACGCCGTCGAAAGAGACAGTCTT
 TCATCTCGCTGCTCGAAACAAAAATATGGATGCCTTTGTTTTTATGGCAGAGAGTTGGG
 AATTAACAGCCAAATTCTTCTACAGCAACCGATGAAAGTGGCAACACTGTCTTACATAT
 TGCTGCATCCGTCTCTTTGATGCTCCTCTTATACGTTACATTGTTGGTAAGAATATAGT
 AGATATCACGTCCAAGAACAAGATGGGTTTGAAGCTTTTCAACTTCTCCCTCGAGAAGC
 CCAAGACTTTGAGTTGTATCAAGGTGGCTGAGATTGGTACCGAGACTTCACAAGAGCT
 GGATTCTGAGAACAATGTAGAACAACACGAAGGCTCTCAAGAGGTCGAGGTAATACGGTT
 GCTAAGGATTATAGGAATAAACACATCAGAGATAGCAGAGAGAAAGAGAAGCAAGGAACA
 GGAAGTGGAAAGAGGTCGTGAGAACTTGAATATCAGATGCATATAGAAGCATTACAGAA
 TGCAAGAAATACGATTGCTATAGTGGCAGTCTTGATTGCTTCAGTTGCTTATGCCGGTGG
 GATAAACCCCTCCGGGGGGCGTCTACCAAGACGGGCCATGGAGAGGGAAATCCTTAGTGGG
 GAAAACAACGGCGTTTAAGGCTTTGCGATATGCAACAACATCGCACTGTTACGTCCTT
 GGGCATCGTTATTCTTCTTGTAGCATCATACCTTACAAGAGGAAACCCTTAAAGAGATT
 ATTGGTGGCCACGCATAGGATGATGTGGGTTTCTGTAGGTTTCATGGCGACGGCTTATAT
 AGCGGCGTCTTGGGTGACCATAACCGATTATCATGGAACACAATGGTTATTTCCAGCAAT
 TGTAGCCGTTGCTGGTGGAGCGTTGACCGTACTCTTTTCTATCTCGGAGTTGAGACCAT
 CGGTCAATGGTTTAAAGAAGATGAATCGTGTAGGGGATAATATACCTTCCTTTGCAAGAAC
 CAGTTCAGATTTAGCCGCTCCCGGAAAATCAGGCTATTTACCTATTAAGAAAACTGGT
 TTTCTAATTTCCCTGTAACCTGTGTAATTGTGTATGTG

>G2032 Amino Acid Sequence (domain in AA coordinates: entire protein)

MQPIFHAILKNDLPFALELVEDSESSLEERNEEEHLNNTVLHMAAKFGHRELVSIIELR
 PSLVSRNAYRNTPLHLAAILGDVNIVMQMLETGLEVCARNINNHPLHLACRSNSIEA
 ARLIAEKTQSIGLGELILAISSGSTSIVGTILERFPDLAREEAWVVEDGSQSTLLHHACD
 KGDFELTTILLGLDQGLEALNPNGLSPLHLAVLRGSVVILEEFLDKVPLSFSSITPSKE
 TVFHLLAARNKNMDAFVFMESLGINSQILLQQTDESGNTVLHIAASVSFDAPLIRYIVGK
 NIVDITSKNKMGFEAFQLLPREAQDFELLSRWLRFGTETSQELDSENNVEQHEGSQEVFV
 IRLRLRIIGINTSEIAERKRSKEQEVERGRONLEYQMHEALQNARNTIAIVAVLIASVAY
 AGGINPPGGVYQDGPWRGKSLVGKTTAFKVFAICNNIALFTSLGIVILLVSIIPYKRKPL
 KRLLVATHRMMWVSVGFMATAYIAASWVTIPHYHGTQWLFPPIAVAVAGGALTVLFFYLGV
 ETIGHWFKKMNRVGDNIPTSSDLAVSGKSGYFTY*

>G1396 (83..313)

TCGACCTCGTTTCCTTCTCCTCTCTTCCCTACCATTAGTACGTTACTGGAGCTGATCTC
 ACGTATATTTGGATCGTAATCATGGACGGCGAAGATTTGCCGGAAGCGGCTGCTGA
 AGCCAAGGGATGAACCCGGGATTAATCGTGTCTGCTTGTGTTGGAGGTCGCTTCTTGT
 GTTCTTAATCGCAACTACGTGCTTTACGTTTATGCTCAGAAGAACCCTACCTCCAAGGAA
 GAAGAAGCCCGTTTCAAAAAGAAGCTCAAGCGGGAGAAGCTAAAGCAAGGAGTCCCTGT
 CCCTGGAGAATAAAAGCCAGCTTAAGCTTCTTCACTTGTGCCTCCTTCAAAGCGGTTTT
 TGTTGGTTACCAAAATTTACCTTGCAGGTTTTTTTTCTTCTTTACTTCTGTCATGAGG
 ATTATCTTTGAGGCCT

>G1396 Amino Acid Sequence (domain in AA coordinates: TBD)

MDGEDFAGKAAAEAKGLNPGLIVLLVVGGLLVFLIANVLYVYAQKNLPPRKKKPVSKK
 KLKREKLKQGVVPVGE*

>G619 (382..2748)

ATTTTTTCCAACTGCAAATTTTAGTCTATGTCTGTTCCCTTGTGCTCCCTCTTCTCAGT

>G619 Amino Acid Sequence (domain in AA coordinates: 64-406)
MEFDLNTIEIAVEVEEENDDVGVGVGGGTRIDKRLGISPSSSSSSSCSSGSSSSSSSTGSAS
SIYSELWHACAGPLTCLPKKGNVVVYFPQGHLEQDAMVSYSSPLEIPKFDLPNQIVCRVV
NVQLLANKDITDEVYTVTLLPLQEFMSMLNGEGKEVKELGEEERNSSSVKRTPHMFCKT
LTASDTSTHGGKSVPRRAEDCFAPLDEYKQQRPSGLLEIAKDLHGVEWKFRIYRGQPRRH
LLTTGWSIFVSGKNLVSGDAVLFLRDEGVEGLRQIRRAAPRNGLPDSIIEKNSCSNLS
LVANAVSTKSMFHVYSPRATHAEFVPIYEKYITSIRSPVCIGTRFRMRFEEMDSPPERRC

AGVVTGVCDDLDPYRWPNKWRCLLVWRWDESFVSDHQERVSPWEIDPSVSLPHLSIQSSPR
PKRPWAGLLDTPPGNPITKRGGLDFEESVRPSKVLQEQENIGSASPSQGFDMNRIL
DFAMQSHANPVLVSRVKDRFGEFVDATGVNPACSGVMDLDRFPRVLQGEICSLKSPQ
FAGFSPAAAPNPFAYQANKSSYYPLALHGIRSTHVPYQNPYNAGNQSSGPPSRAINFGEE
TRKFDAQNEGGLPNNVTADLPFKIDMMGKQKGSELNMNASSGCKLFGFSLPVETPASKPQ
SSSKRICTKVHKQGSQVGRAIDLRLNGYDDLMELELRFNMEGLLRDPEKGWRILYTD
ENDMMVVGDDPWHDPCNVVWKIHLTYKKEVENANDDNKSCLEQAALMMEASKSSSVSQPD
SSPTITRV*

>G2295 (33..917)

GTAATATATAACAATAACTCAGGTTACAAAGGATGGTTCCGAAAGTGGTCGACCTACAAA
GGATAGCGAACGATAAGACAAGGATAACAACCTACAAGAAGAGGAAAGCTAGTCTTTACA
AGAAGGCACAAGAGTTCTCAACTCTCTGCGGCGTCGAGACATGTCTCATCGTCTACGGTC
CCACGAAGGCTACCGATGTGGTGATTTCCGAGCCAGAGATATGGCCGAAGGACGAGACCA
AAGTCAGGGCCATCATACGCAAGTACAAAGACACAGTGTGACCGAGCTGCAGGAAAGAAA
CCAACGTGGAGACTTTTCGTCAACGATGTAGGGAAAGGAAACGAGGTGGTGACTAAAAAGA
GAGTGAAGCGTGAGAATAAGTATTCTAGTTGGGAGGAGAAGCTAGACAAGTGTTACGAG
AGCAACTACATGGGATTTCTGTGCCGTGGATAGCAAGTTAATGAAGCTGTAACGAGAC
AGGAGCGTAGTATGTTTAGGGTTAATCATCAAGCCATGGACACACCATTTCCCGCAGAATT
TAATGGACCAACAATTTCATGCCACAGTATTTTCATGAGCAGCCACAGTTTCAAGGCTTCC
CTAATAATTTCAATAATATGGGTTTCTCGTTGATTTACCTCATGATGGTCAGATTCAAA
TGGACCCAAATCTCATGGAGAAGTGGACCGACTTGGCTTTGACTCAAAGCTTGATGATGT
CAAAGGGAAACGATGGTACTCAATTCATGCAGAGGCAAGAACAACCATACTATAATCGTG
AACAGGTGTATCGAGGTCTGCAGGTTTCAATGTTAACCCTTTATGGGATATCAAGTCC
CGTTTAAATATTCCTAATTGGAGATTATCGGGAAATCAAGTTGAAAATTGGGAGCTTTCAG
GGAAGAAAACGATATGATTGAATTACGGAGCTTTATTAGTTTTTAGGGTTTTATAGTTT
TG

>G2295 Amino Acid Sequence (domain in AA coordinates: TBD)

MVPKVVDLQRIANDKTRITTYKKRKASLYKKAQEFSTLCGVETCLIVYGPTKATDVVISE
PEIWPKDETKVRAIRKYKDTVSTSCRKETNVETVNDVVGKNEVVTKRVKRENYSSW
EEKLDKCSREQLHGIFCAVDSKLEAVTRQERSMFRVNHQAMDTFPQNLMDQQFMPQYF
HEQPQFGFPNNFNMMGFSLISPHDQGIQMDPNLMEKWTDLALTQSLMMSKGNDDGTQFMQ
RQEQPYNREQVVSRSAGFNVNPFMGYQVPFNIIPNWRLSGNQVENWELSGKKTII*

>G312 (1..1755)

ATGGCTTACATGTGCACTGATAGTGGCAATCTAATGGCTATTGCTCAACAAGTCATCAAA
CAGAAGCAGCAACAAGAACAACAACAGCAGCAACATCATCAAGACCATCAGATTTTTGGT
ATTAATCCTTTGTCTCTTAAACCCATGGCCCAATACTTCCCTCGGGTTTGGGCTTTCAGGT
TCGGCTTTTCCCGACCGGTTTCAAGTTACCGGCGGCGGAGATTCCAACGATCCTGGCTTT
CCTTTTCTTAACCTTAGACCACCAACACGCCACAACCACCGGCGGTGGGTTTCAGGTATCT
GATTTTCGGCGGTGAACCGGCGGCGGCGAGTTTGAATCCGACGAGTGGATGGAGACTCTT
ATCAGCGGTGGAGACTCCGTTGCAGACGGTCCCTGATTGTGACACCTGGCATGATAATCCC
GATTACGTAATCTACGGTCCCTGATCCATTTCGATACTTACCCGAGTCGACTCAGTGTCCAA
CCGTTCAGATCTAAACCGAGTCATTGACACGTCGAGTCCGCTTCTCCGCCGACCTTGTGG
CCTCCTTCTTCGCCATTATCGATTCTCCTCGCTTACTCATGAGTCACCAACCAAGAAGAT
CCAGAGACTAACGACTCCGAAGACGATGACTTCGACCTAGAACCACCTCTCCTCAAAGCT
ATATACGACTGTGCACGGATCTCAGACTCTGACCCTAACGAAGCTTCCAAGACGCTTCTT
CAGATCCGAGAATCTGTATCGGAGCTAGGTGATCCGACGGAGCGAGTTGCATTTTACTTC
ACGGAAGCTCTCTCAACAGACTGTCTCCTAATTCGCCGCGGACGTCGTCTTCTTCTTCA
TCTACGGAGGATTTAATCTTATCTTATAAAACCCTAAACGACGCTTGTCTTACTCCAAA
TTCGCACATTTGACGGCGAATCAAGCGATTCTAGAAGCGACGGAGAAGTCGAACAAGATT
CACATCGTCGATTTTGAATCGTTCAAGGTATACAATGGCCTGCTCTTCTTCAAGCTCTA
GCTACTCGTACTTCTGGTAAACCCACTCAAATCCGGTCTCGGGTATACCCGCTCCATCT
CTCGGTGAATCTCCGGAACCGTCTGTTAATCGCCACCGGAAACCGCTCCGTGATTTCCGC
AAGGTTCTGGATCTGAATTTGATTTTCATCCCAATTCCTACTCCCATACATTTACTTAAC
GGGTCAAGTTTCCGGGTGACCCGGATGAAGTACTGGCCGTGAATTTTCATGCTCCAGCTC
TACAAATTACTCGACGAGACGCCGACGATAGTTGACACCGCACTACGGCTCGCCAAATCG
TTGAACCCGAGGGTCTGCTACTCTCGGAGAATACGAAGTGAGCTTAAACCGGGTCCGTTTC
GCTAACCCGGTAAAGAACCGCTTCAATTCTATTCCGCGGTTTTTGAATCCCTTGAACCG

AACTTGGGGCGTGATTTCGGAGGAGAGAGTGAGAGTTGAGCGAGAGTTGTTTCGGCCGGAGA
 ATCTCGGGTTTGATTGGACCGGAGAAAACCGGAATTCATAGAGAAAGAATGGAAGAGAAA
 GAGCAATGGCGGGTATTAATGGAGAATGCCGGTTTTGAATCGGTTAAGCTGAGTAATTAC
 GCAGTGAGCCAAGCGAAGATTCTATTGTGGAATTACAATTACAGCAATTTGTATTCAATT
 GTTGAATCTAAGCCTGGCTTCATCTCTTTGGCCTGGAACGATTTACCTCTCCTCACTCTT
 TCTTCTGGCGATAA

>G312 Amino Acid Sequence (domain in AA coordinates: 320-336)
 MAYMCTDSGNLMAIAQQVIKQKQQEQQQQHHQDHQIFGINPLSLNPWPNTSLGFGLSG
 SAFDPDFQVTGGGDSNDPGFPFPLNDHHHATTTGGGFRLSDFGGGTGGGEFESDEWMETL
 ISGGDSVADGPDCTDWHNDPDYVIYGPDPFDTPSRLSVQPSDLNRVIDTSSPLPPPTLW
 PPSSPLSIPPLTHESPTKEDPETNDSDDDDFDLEPPLLKAIYDCARISDSDPNEASKTLL
 QIRESVSELGDPTEFVAFYFTEALSNNLSPNSPATSSSSSSSTEDLILSYKTLNDACPYSK
 FAHLTANQAILEATEKSNKIHIVDFGIVQGIQWPALLQALATRTSGKPTQIRVSGIPAPS
 LGESPEPSLIATGNRLRDFAKVLDLNFDFIPILPTIHLNNGSSFRVDPDEVLA VNFMLQL
 YKLLDEPTIVDTALRLAKSLNPRVVTLGEYEVSLNRVGFANRVKNALQFYSAVFESLEP
 NLGRDSEERVVRERELFGRRISGLIGPEKTGIHRERMEEKEQWRVLMENAGFESVKLSNY
 AVSQAKILLWNYNYSNLYSIVESKPGFISLAWNDLPLLTLSSWR*

>G1444 (192..1001)

AATCCCCTATCCTTCGCAAGACCCTTCCTCTATATAAGGAAGTTCATTTTCAATTTTTTTT
 GACACGCTGACAAGCTGACTCTAGCATATCTGGCACCGGCGACAGTCTCTTTGGTGTC
 AAAGATCCCCAAAATCAAATCGAAAGAGAGATAAATCAAAGGAAGAATCTTTATCT
 GCTTCTCTCGATGAGGATCCGGAACGACAAGTGCCTCTTCCTTTATCGTCTCTATTAC
 CAGTTCCTCTATCAGATCTCTACTTTAACCGCTCACCGACGGCCACCGCGAGATACTTC
 GCGGTGGTTATAAAGACGGCGGTGATGATTTTGGTTCTCTTCAGCTTTCGCTTCGCCGC
 CGTCGAGATTTCTGATCGGCTTATTCAAAGAGATTTGATAAAGAAGAAGGAGGAGTCA
 AGCCTTTGGATGATGATAATGGTGATGTAGACGTCAAGAGTCGTACTGATGCATCGGGCA
 GCAAGAATGTTAATCCCCGAGGAGAATCCGTCTCTTCAATACAAGTTGTGAGAAGAATG
 AAAAGGTTGTGTCTTTGAGGAAGAGAAGAGGCTTTATCAACTTTGAGGATTACGAAGATG
 AGGAAGATGAAGAAGCTAGTGGCGGTGGAGGCCGTATTAATAAAGGGAAAAAGAAAGCGA
 AAAAGAGCGGTGGTGGGTAGAGGAAGGATCACGGTGCAGCCGTGTTAACGGTAGAGGAT
 GGAGATGTTGTGACGAAACGCTTGTGGTTATTCTCTTTGTGAGCATCATCTCGGTAAAG
 GAAGGGTAAGGAGCATGAACAAGAGTGGTGGTGGTGGTGGTGGCGGAGAAAGAGCGGTGG
 TGGTGGAAAGTGAAGAAGAAGAGAGTAAAGCTTGGCATGGTAAAGGCACGTTCAATAAGTA
 GTTTGCTTGGACAAACCAGCACTAGTGGTGGTACTAGTGGTGATGTTGATCAGGGTGAGA
 TAAGTGACCTGCTGATCAGTTTCGTGATGTGATAAGTAGGTCTGTTGATCAGCATTTG
 CATGTATATGGATATGTGTATGTTTATGTATCATGATGATAATGGGCATAGCGCGGCCGCT
 CTAGACAGGCCCTGGAACCGGATCCTCTAGCTAGAGCTTTCGTTAGTATCATCGGGTTAG
 ACAACGTT

>G1444 Amino Acid Sequence (domain in AA coordinates: 168-193)
 MRIRKRQVPLPLSSLLPVPLSDLYFNRSPTATARYFRGGYKDGDDDFGSLQLSLPPPSQI
 SDRLIQRDLIKKKEEVKALDDNDGDVDVKSRTDASGSKNVNPRGESVSSIQVVEKNEKVV
 SLRKRRGFINFEDYEDEEDEEASGGGRINKGKKKAKKSGGGLEEGSRCSRVNGRGWRCC
 QQTLVGYSLCEHHLGKGRVRSNMKSGGGRGGEKKAVVVEVKKRVKLGVMKARSISLLG
 QTSTSGGTSGDQGEISAPADQFAACDK*

>G801 (27..746)

GATAGTGATAACGAAATCCTAATTCATGGCCGACAACGACGAGCAGTGAGTAACGGCA
 TCATAGTCGAGCAGACGTCAAACAAAGGACCTCTTAACGCCGTTAAGAAACACCGTCTA
 AAGATCGACACAGCAAAGTTGACGGAAGAGGAAGAAGGATTCGTATGCCAATCATTTGCG
 CAGCTCGAGTTTTTCAATTGACCAGAGAGTTAGGTCAAAAGTCCGATGGTCAAACCATAG
 AGTGGCTTCTCCGTCAAGCTGAGCCTTCTATCATAGCCGCCACTGGAACCTGGCACTACTC
 CGGCGAGTTTCTCCACTGCTTCTCTCTCCACTTCTTCTCCGTTTACTCTCGGGAAACGTG
 TCGTCAGAGCGGAGGAAGGAGAATCCGGCGCGGAGGAGGAGGGTTAACAGTGGGAC
 ACACAATGGGGACTTCGTTAATGGGTGGTGGTGGTTCGGTGGGTTTTGGGCTGTTCCGG
 CGAGGCCGATTTCCGACAAGTCTGGAGCTTTGCAACCGGAGCTCCACCGGAAATGGTTT
 TTGCGCAGCAGCAGCAACAGCTACACTCTTCGTCGCCACAGCAGCAACAGCAAGCTT
 CCGCCGCCGAGCAGCTGCAATGGGTGAGGCTTCAGCAGCTAGAGTTGGGAATTATCTTC
 CGGGTCATCATCTCAATTTGCTTCTTTGTCTGGTGGAGCTAACGGGTGGGTTCGGA

GGGAAGACGACCACGAACCACGTTGAGAAATGGTATTGTCTTTTGGTAATGTATAGAAA
AATTCCTATGTTTATGTCATCGAAAGTGTGTTAGAAAGTACCTCTAATTTGCGGTTTCTT
TTGCTCCTTTTACTTAATTTAAGCTTATTGCTTGTTTGATTAGGGTTTTAGGGTTTAA
GAATATTTGGTCTCGTTAATTTGTTTCGGAGAGTGATAGAAAAGAGAGAGATTGATTGA
TTGTTGTACCTAAAACGCTATAAAAGCTCTGTTTTTACTAGCGAAAAAA

>G801 Amino Acid Sequence (domain in AA coordinates: 32-93)
MADNDGAVSNGIIVEQTSNKGPLNAVKKPPSKDRHSKVDGRGRRIRMPIICAARVFQLTR
ELGHKSDGQTIEWLLRQAEPSIIAATGTGTPASFSTASLSTSSPFTLGKRVVRAEEGES
GGGGGGGLTVGHMTGSLMGGGGSGGFVAVPARPDFGQVWSFATGAPPEMVFAQQQQPAT
LFVRHQQQQASAAAAAAMGEASAARVGNYPGHHLLLASLSGGANGSGRREDDHEPR*
>G1950 (42..764)

CTGAATTCGAACTTTGGAAGAAGAAGCTTTGATCAATCATGGAAATTGCAACCGATA
CAGCAAAGCAGATGAGAGACGAAGAGTTGTTCAAAGCAGCGGAATGGGGAGATTTCATCGT
TGTTTCATGTCATTATCTGAAGAACAGCTCTCTAAATCTCTCAATTTAGAAAACGAAGATG
GTCGCTCTCTCCTCCATGTGCTGCTTCTCGGCCATTCTCAAATAGTGAAGTTGTTAT
CAAGTTTCAGATGAAGCAAAGACTGTAATCAATAGCAAGGATGATGAAGGATGGGCTCCTT
TGCATTCCGCTGCTAGCATCGGTAATGCTGAGCTCGTTGAGGTGCTTTTGACCAGAGGTG
CTGATGTCAATGCCAAAAATAACGGTGGTTCGCACTGCTCTTCACTATGCTGCTAGCAAAG
GCCGGTTGGAGATTGCTCAGCTTTTATTAACACACGGTGCAAAGATTAACATCACAGACA
AGGTTGGTTGCACTCCGCTTCACAGGGCAGCAAGCGTGGGAAAGTTAGAAGTTTGTGAAT
TTCTTATTGAAGAAGGAGCAGAGATCGATGCTACGGATAAAATGGGTCAAACCTGCACTCA
TGCATTCACTTATCTGCGATGACAAACAGGTTGCGTTCCTGCTTATAAGACATGGTGCAG
ATGTGGATGTAGAAGACAAGGAAGGCTACACTGTTCTAGGCCGAGCTACCAATGAATTCC
GACCTGCACTTATCGATGCTGCTAAGGCCATGCTTGAAGGATAAAATGACTCTGGATTAC
TTTAAACCTTACTAATCTGAGAGTTGTTTAGTTACTTAAAGGATTTTCTTTACTGTA
TCATGTTTGCAAAATGTTTCTGCCTTATCAATTCATGTTCTGT

>G1950 Amino Acid Sequence (domain in AA coordinates: 65-228)
MEIATDTAKQMRDEELFKAAEWGDSSLFMSLSEQLSKSLNFRNEDGRSLLHVAASFGHS
QIVKLLSSSDEAKTVINSKDDEGWAPLHSAASIGNAELVEVLLTRGADVNAKNNGGRTAL
HYAASKGRLEIAQLLLTHGAKINITDKVGCTPLHRAASVGKLEVCEFLIEEGAEIDATDK
MGQTALMHSVICDDKQVAFLLIRHGADVDEDKEGYTVLGRATNEFRPALIDAAKAMLEG
*

>G958 (55..1950)
CGTCGACATGTTTCATATTGTTTCTAGCTAAGAAGTTTGTATAAGGCAGTGGACATGGCT
CCTGTTTCAATGCCTCCAGGTTTCCGGTTTTCATCCAACAGACGAAGAGCTTGTATATAC
TACCTCAAGCGAAAGATTAATGGTCCGACTATTGAGTTAGAGATAATACCCGAGATTGAT
CTTTACAAATGCGAACCTTGGGATTTACCTGGGAAGTCCTTGCTGCCAAGTAAAGACCTA
GAATGTTCTTTTTTTCAGTCTTCGAGACCGGAAATATCCAAACGGATCAAGAACAAACCGG
GCGACCAAAGCAGGTTACTGGAAAGCCACCGGGAAGATCGTAAAGTGAATTCACATTCA
CGGATGGTTGGAACAAAGAAAACATTAGTTTATTACCGAGGAAGAGCGCTCATGGCTCT
CGTACCGATTGGGTATGCACGAGTACCGTCTTGAAGAACAAGAATGTGACTCTAAATCC
GGTATACAGGATGCCATGCACTTTGTCGAGTATTTAAGAAGAGTGCTTTAGCCAACAAA
ATTGAAGAACAACACCATGGTACGAAGAAGAACAAGGAACGACTAATAGTGAACAATCT
ACTTCTAGTACTTGTGTTGTTATCTGATGGAATGTATGAAAACCTCGAAAACCTCGGGGTAT
CCAGTCTCACCTGAGACAGGAGGCTTAACCTCAACTCGGTAATAATTCTGTCGTCGGATATG
GAAACGATAGAGAATAAATGGAGTCAGTTTATGTCGCATGACACGTCCTTCAACTTCCCA
CCTCAGTCTCAATATGGAACAATCTCATATCCTCCCTCGAAGGTTGATATAGCGTTAGAG
TGTGCAAGACTACAAATCGTATGTTGCCACCAGTACCACCCTTTACGTAGAAGGTCTC
ACACACAATGAATATTTTGGAAACAATGTAGCTAACGATACAGATGAAATGTTGAGCAAG
ATTATAGCATTGGCTCAAGCCTCACATGAGCCACGAAACAGTCTAGACTCATGGGACGGT
GGTTCTGCTTCCGGGAACCTTCATGGAGACTTTAATCTTCCGGAGAAAAAGTCTCATGC
CTAGAGGCGAACGTGGAGGCTGTAGATATGCAAGAACACCATGTGAATTTTAAAGGAAGAA
AGACTTGTGAAAACCTTGAGATGGGTAGGAGTATCAAGCAAGGAACCTGAAAAGAGCTTC
GTTGAAGAACAACCTCAACGGTAATTCCTATAGAAGATATTTGGAGATATCATAATGATAAT
CAAGAACAAGAACATCATGATCAAGATGGTATGGACGTTAACAACAACAATGGAGATGTG
GATGATGCTTTTCACTCGAGTTTTCGGAAAACGAACATAACGAGAATCTTTTGGACAAG
AACGATCATGAGACAACGAGTTCCTCATGTTTTGAGGTGGTAAAAAAGTTGAGGTTAGC

CATGGATTGTTTGTACAACTCGTCAGGTAACCAACACATTCTTCCAACAGATAGTACCA
 TCGCAAACCGTTATAGTTTATATAAATCCGACGGATGGCAATGAGTGTTCATAGTATG
 ACATCAAAAGAGGAGGTTTCATGTCCGTAAAAAGATAAATCCGCGAATCAACGGAGTAAGC
 TCAACAGTTCTTGGACAATGGAGAAAATTCGCGCATGTTATTGGCTTCATTCCTATGCTT
 CTATTGATGCGTTGTGTTTCATCGAGGTAACCTAACAAAAACAGAGGCAGTGAAGGTTAC
 TCGAGGCAGCCTACGAGAGGAGATTGTAACAATCGGGGAACAATACTCATGATGGAAAAT
 GCTGTCTGAGAAGAAAAATTTGGAAGAAGAAGAAAGAGAAAAATATGGTTGACGAACAA
 GGTTTTCGGTTTCAAGATAGTTTCGTATTGAAGAAGTTGGGGCTTTCTCTTGCTATCATC
 TTAGCTGTTTCTACCATAAGTCTTATTTGAATACTGAGGTTCAATATATCATATATGGCT
 TTTCACTTTTCTATTGTACTCCCATTTGCCTAGGTCGTATGC

>G958 Amino Acid Sequence (conserved domain in AA coordinates: 7-156)
 MAPVSMPPGFRFHPDEELVIYYLKRKINGRTIELEIIPIDLYKCEPWLDPGKSLLP
 DLEWFFFSRDRKYPNGSRNTRATKAGYWKATGKDRKVTSHSRMVGTGKTLVYYRGRAPH
 GSRTDWMHEYRLEBQBCDSKSGIQDAYALCRVFKKSALANKIEEQHHGTTKKNKGTNNSE
 QSTSSTCLYSDGMYENLENSGYPVSPETGGLTQLGNSSSDMETIENKWSQFMSHDTSPN
 FPPQSQYGTISYPPSKVDIALECARLQNRMLPPVPLYVEGLTHNEYFGNNVANDTDEML
 SKIIALAQASHEPRNSLDSDWGGASGNFHGDFNYSGEKVSCLANVEAVDMQEHVNF
 EERLVENLRWVGVSSEKELEKSFVEEHSVIPIEDIWRHYHNDNQEQEHDQDGMVNNNG
 DVDDAFTLEFSENEHNENLLDKNDEHTSSSCFEVVKKEVSHGLFVTTTQVNTTFFQOI
 VPSQTVIVYINPTDGNCCSMTSKEEVHVRKKINPRINGVSVTLGQWRKFAHVIGFIP
 MLLLMRCVHRGNSNKNRGSEGYRQPTRGDCNNRGITILMMENAVVRRKIWKKKKEKNMVD
 EQGFRFQDSFVLKLLGLSLAILAVSTISLI*

>G1037 (1..1722)

ATGACTGTTGAACAAAATTTAGAAGCTTTGGATCAGTTTCCTGTAGGAATGAGAGTTCTT
 GCTGTTGATGATGACCAAACTTGTCTCAAAATCCTTGAATCTCTCCTTCGTCACTGCCAA
 TACCATGTAAACACGACGAACCAAGCACAAAAGGCTTTAGAGTTATTGAGAGAGAACAAAG
 AACAAGTTTGATCTGGTTATTAGTGATGTTGACATGCCTGACATGGATGGTTTCAAACCT
 CTTGAGCTTGTGGTCTTGAAATGGACCTACCTGTCTATAATGTTGTCTGCGCATAGTGAT
 CCAAAGTATGTGATGAAGGAGTTACTCATGGTGCTTGTGATTATCTACTGAAGCCGGTT
 CGTATTGAGGAGTTGAAGAACATATGGCAACATGTCGTGAGAAGTAGATTGATAAGAAC
 CGTGGGAGTAATAATAATGGTGATAAGAGAGATGGATCAGGTAATGAAGGTGTTGGGAAT
 TCTGATCCGAACAATGGGAAAGGTAATAGAAAACGTAAAGATCAGTATAATGAAGATGAG
 GATGAGGATAGAGATGATAATGATGATTCTGTGCTCAAAAGAAGCAACGTGTTGTTGG
 ACTGTTGAGCTGCATAAGAAATTTGTTGCAGCTGTTAACCAATTGGGATATGAGAAGGCT
 ATGCCTAAAAGATTTTGGATCTGATGAATGTTGAGAAGCTCACTAGAGAAAATGTGGCC
 AGTCATCTTCAGAAATTCGCCTTTACTTGAAGAGGATCAGTGGTGTGGCTAATCAGCAA
 GCTATTATGGCAAACTCTGAGTTACATTTTATGCAAATGAATGGACTTGATGGTTTCCAT
 CACCGCCCAATCCCTGTTGGATCTGGTCAGTACCATGGTGGGGCTCTGCAATGAGATCT
 TTCCCTCCAAACGGGATTCTTGGCAGACTCAATAGCTCTTCGGGGATCGGTGTCCGCAGC
 CTTTCTTCTCCTCTGCAAGGAATGTTCTTGCAAAACAGACCGATATCGGAAAGTTTCAC
 CATGTCTCATCACTTCTCTTAACCACAGTGATGGAGGAAACATCTTCAAGGGTTGCCA
 ATGCCTTTAGAGTTGACACAGCTTCAGACAAACAACAACAAAGTAGAAACATGAACAGT
 AACAAGAGCATTGCTGGGACCTCCATGGCTTTTCTAGCTTCTCTACGCAACAAACTCG
 CTCATCAGTGCTCCTAATAACAATGTCGTGGTTCTAGAAGGTCAACCAAGCAACTCCT
 CCAGGCTTCCAGGACACAGATCAATAACGTTTGGAGCATTTGGTCAAATGCTGTATCC
 TCTTCGACTCAACCTCCTCCCCGGCACATAACAGTAATAGTATCAATCATCAGTTTCGAT
 GTCTCTCCATTACCGCATTCTAGACCCGACCCCTTGGAAATGGAACAATGTGTATCAAGC
 TACTCTATACCATTTCTGTGACTCTGCCAATAACATTGAGTTCTCCAGCCTTGGATACAACA
 AATCCCCGAGCTTTCTGTAGAAACACGGACTTCGATTCAAACACAAATGTGCAACCTGGA
 GTCTTTTATGGTCCATCCACGGATGCTATGGCTCTGTTGAGTAGTAGTAACCCGAAAGAA
 GGGTTCGTCGTAGGCCAACAGAGTTACAGAGTGGTGGATTCTATGGTTGCAGATGCTGGT
 TCCTTAGATGATATAGTCAACTCCACGATGAAGCAGGTGTA

>G1037 Amino Acid Sequence (domain in AA coordinates: 11-134, 200-248)
 MTVEQNLEALDQFPVGMRLAVDDDDQCLKILESLLRHCQYHVTTTNAQKALELLRENK
 NKFDLVISDVMDPMDGFKLLELVGLEMDLPVIMLSAHSDPKYVMKGVTHGACDYLLKPV
 RIEELKNIWQHVVRSRFDKNRGSNNNGDKRDGSGNEGVSNDPNNKGKRNKRKDQYNEDE
 DEDRDDNDDSCAQKKQVRVVTVELHKKFVAAVNQLGYEKAMPKKILDLMNVEKLTRENV

SHLQKFRLLYLRISGVANQQAIMANSELHFMQMNGLDGFHHRPIPVGSGQYHGGAPAMRS
FPFNGILGRNLNSSSGIGVRSLSPPAGMFLQNQTDIGKFHHVSSLPLNHSDDGNNILQGLP
MPLEFDQLQTNNNKSRNMNSNKSIAGTSMAPFSFSTQQNSLISAPNNNVVLEGHPOATP
PGFPGHQINKRLEHWSNAVSSSTHPPPAHNSNSINHQFDVSLPHSRPDPLEWNNVSSS
YSIPFCDSSANTLSSPALDTTNPRAFRCNTDFDSNTNVQPGVFGPSTDMALLSSSNPKE
GFVVGGQKLQSGGFMVADAGSLDDIVNSTMKQV*

>G2065 (33..1124)

AACCACACAAAACAAAACAAAAAACATATTGATGGGGATGAAGAAGGTAAAGCTATCTT
TGATAGCTAATGAAAGATCAAGGAAAACATCCTTCATGAAGAGGAAAAACGGGATATTCA
AGAACTCCACGAGTTGTCAACTCTATGTGGTGTCCAAGCTTGTGCTCTCATCTATAGTC
CATTATACCGGTTCCAGAGTCATGGCCGTCAAGGGAAGGTGCTAAAAAGGTAGCTTCAA
AGTTTCTGGAGATGCCCGGACAGCCCGAACCAGGAAGATGATGGATCAAGAAACCCATC
TTATGGAGAGGATTACCAAAGCAAAAGAGCAACTAAAGAATTTGGCTGCTGAGAACCGAG
AATTACAGGTTAGACGATTATGTTTGTATTGTGTTGAAGGCAAAATGTCCAGTATCGTT
ATGATGCAAAAGACCTTCAAGATTTGCTATCTTGTATGAATCTATATCTCGATCAGCTTA
ACGGAAGGATCGAGTCCATTAAAGAAAACGGTGAGTCGTTGTTGTCTTCCGTCTCTCCTT
TTCCTACTAGAATTGGTGTGACGAAATTGGTGATGAGTCGTTTTCCGACTCTCCTATTC
ATTCTACAACCTAGGGTTGTAGATACTCCTAATGCTACCAATCCTCATGTTCTTGCGGGCG
ATATGACTCCTTTTCTTGATGCGGACGCAAATGCGGTAAGTCTCCAGTCGATTTTCTG
ATCATATTCAATATGAAAATATGAATATGAGTCAAAATCTGCATGAACCGTTTCAACACC
TTGTTCTCTACTAACGTTTGTGATTTTATCAAAATCAGAATATGAATCAGGTTCAATACC
AGGCTCCTAATAATCTGTTTAAATCAGATTCAACGAGAATTCTACAACATAAATTTGAATC
TGAATTTGAATCTGAATTCAAATCAGTATCTGAATCAACAACAATCATTATGAATCCGA
TGGTGGAACAACATATGAATCATGTTGGAGGGCGTGAAAGCATTCTTTCTGAGACAGAA
ACTACTACAACATACTAATCAACTACCAGCCGTTGATCTTGCTTCCACCAGTTACATGCCTT
CAACCACCGATGTTTATGATCCTTACATCAACAACAATCTCTAATCAGAAAGACGGAGA
TTTTCTAGTTTAA

>G2065 Amino Acid Sequence (domain in AA coordinates: TBD)

MGMKKVKLSLIANERSRKTSFMKRNKIGFKKLHELSTLCGVQACALIYSPFIPVPESWPS
REGAKKVASKFLEMPRTARTRKMDQETHLMERITKAKEQLKNLAAENRELQVRRFMFDC
VEGKMSQYRYDAKDLQDLLSCMNLVLDQLNGRIESIKENGESLLSSVSPFPTRIGVDEIG
DESFSDSPIHSTTRVVDTPNATNPHVLADMTPLDADANAVTAPSRFSDHIQYENMMS
QNLHEPFQHLVPTNVCDYFQONQNMNQVQYQAPNNLNFNQIQREFYNINLNLNLNSNQYL
NQQQSFMNPMVEQHMNHVGGRESIPFVDRNYNYNQLPAVDLASTSYMPSTTDVDPYIN
NNL*

>G2137 (77..1123)

GGGATTTGACTTTTAGCACTTCAAATCCAAAGCTAAAAGACAAAAAGAATAGAGGTTTCG
ATTTGCATCTCCATTAATGGGCATCGATCTTTCTCTTAAGCTCGAGGCCGAGGAGAAAA
GAAAGAGATAGAAGGATCGAAACATAGCCGTGAGAACAAGAAGACGAAGAACATGATGC
TAGTGGTGTGATGAAGATGAACAAATGGTGAAAGAAGACGAAGATGATCTTCTTTTAGG
TTTAAGAACCCGAGAAGAAGAAAACGAACGTGAAGAGCTCTTGCAGCTACAGATCCAGAT
GGAAAGTGTGAAAGAAGAGAATACTAGGTTGAGGAAGCTTGTGAGCAGACTCTTGAAGA
TTATCGTCATCTTGAGATGAAATTCCCGGTTATCGATAAAACCAAGAAGATGGATCTTGA
AATGTTCTTGGAGTACAAGGCAAACGATGTGTGGATATAACAAGTAAGGCTCGGAAAAG
AGGAGCTGAGAGATCTCCGTCAATGGAAAGAGAAATAGGGCTTTTCACTTTCTCTAGAGAA
AAAACAGAAAACAAGAAGAGAGCAAAGAAGCTGTTCACTCTCATCACCAGATACAATAG
TAGCAGCTTAGATATGAATATGCCACGTATCATTTTCATCTTCTCAAGGTAATAGAAAGGC
CAGGGTGTCCGTGAGGGCGAGATGTGAGACCGCAACAATGAATGATGGATGCCAATGGAG
GAAGTACGGTCAGAAAACCGGAAAGGGAATCCATGTCTCGAGCTTATTACCGATGCAC
CGTGGCTCCAGGATGTCCCGTTAGAAAACAGGTGCAAAGGTGTTTGAAGACATGTCAAT
ACTGATAACAACCTACGAAGGAACACATAACCATCCACTTCCGGTCCGAGCAACAGCCAT
GGCTTCCACTGCCTCTACTTCTCCATTCTTGTACTCGATTCCAGTGACAACCTCTCTCA
TCCTTCCATTATACCAAACCTCAAGCCATAGACTCTTCTTTGATTACATACCCACAAAA
TAGCAGCTACAACAATCGAACCATAGAAGCTTGAACCTTGATGGTCCATCTAGAGGAGA
TCACGTTTCATCTTCTCAAACCGATTAAATTGGATGATGTAGAGTTTCTATATCTCTA
TGCTTGTTCTTTGGTCCCATTTATTTGTCATTATGGATTCTTTGCCTTTCTTCTGTTCTC
GTTTCTAACATTTATGTTTCGTATA

>G2137 Amino Acid Sequence (conserved domain in AA coordinates:109-168)

MGIDLSLKLEAEKKKEIEGSKHSRENKEDEEHDSASGEDEDEQMVKEDEDDSSSLGLRTRE
EENEREELLQLIQMESVKEENTRLRLKLEQTLQEDYRHEMKFPVIDKTKKMDLEMLGV
QGKRCVDITSKARKRGAERSPSMEREIGLSLSLEKKQKQESKEAVQSHHQRYNSSSLDM
NMPRIISSSQGNRKARVSVRARCETATMNDGCQWRKYGQKTAKGNPCPRAYYRCTVAPGC
PVRKQVQRCLEDM SILITTYEGTHNHPLPVGATAMASTASTSPFLLLDSSDNLSPSYQ
TPQAIDSSLITYPQNSSYNRTIRSLNFDGPRGDHVSSSQNRLNWMW*

>G746 (1..1311)

ATGGGTGAGGAGTTAGCTGACACAATGAACCTGGATTTGAATCTTGGGCTGGTCTGAG
TCTGATCTCCAACCTGCACCAAACGAGACTGTGAATTTGGCTGATTGGACTAATGACCCG
CCTGAGAGATCTTCTGAAGCTGTGACAAGGATCAGGACTCGGCATAGGACACGGTTCAGA
CAGCTTAATCTCCCGATCCCGGTTCTATCTGAAACCCATACCATGGCTATAGAGCTCAAC
CAGTTGATGGGAAATTCTGTAAATAGAGCTGCTATGCAGACTGGTGAGGGTAGTGAAAGA
GGCAATGAGGATTTGAAAATGTGTGAGAATGGCGATGGAGCCCTTGGGGACGGTGTATTG
GATAAGAAAAGCGGATGTCGAGAAAAGCAGTGGCAGCGACGGTAACTTTTTCGATTGTAAT
ATATGTTTGGATTTGTGGAAGGAGCCGGTCTCACCTGTTGTGGTCATCTTTACTGTTGG
CCTTGTCTGTACCAATGGTTACAAATTTCCGATGCAAAGGAATGTCTGTTTGTAAAGGA
GAGGTGACCTCCAAAACCGTGACACCGATCTATGGACGTGGAACCAAGAGAGAAATT
GAAGAGAGTTTAGATACTAAGGTCCCCATGAGACCACACGCGAGACGATTGAGAGCTTG
AGGAATACAATTCAAAGGTCGCCTTTTACAATACCAATGGAAGAAATGATTAGACGTATA
CAGAATAGGTTTGACAGGGATTCAACCCAGTCCCTGATTTTAGTAACCGAGAGGCATCA
GAAAGAGTCAACGATCGAGCCAATTCGATCCTTAACCGGTTGATGACATCTAGGGGAGTT
AGATCAGAGCAGCAACAGGCTAGTGCTGCAGCAGCAGCATTGTGCGAGCATCAGAGGAT
ATTGATCTAAATCCAAACATGTGCTCCTGATCTTGAAGGAGAAAGCAACACGAGATTCCAT
CCTCTCTTGATCAGGAGACAGTTACAGTCGCACCGAGTTGCAAGGATCTCGACTTCACT
TCTGCGTTGAGTTGAGCTGAGAGGCTTGTGGATGCGTATTTTAGGACTCATCCGTTGGGG
AGGAACCAACAAGAGCAAAACCATCATGCTCCTGTTGTGGTTGATGATAGAGACTCATTC
TCAAGCATTGCAGCTGTTTATAAATCTGAGAGTCAAGTGGATACTGCAGTTGAGATCGAT
TCTATGGCTCTTTTCGACATCGTCTCGAGGAGAAGGAATGAGAATGGTTTCGAGGGTTTCT
GATGTAGACAGTGCAGATTCTCGTCCGCCTAGGAGAAGGAGATTTACTTGA

>G746 Amino Acid Sequence (domain in AA coordinates: 139-178)

MGEELADTMNLDLNLGPGPESDLQAPNETVNLADWINDPPERSSEAVTRIRTRHRTFR
QLNLPPIPVLSETHMAIELNLQMLGNSVNRAAMQTGEGSERGNEDLKMCENGDGLDGLV
DKKADVEKSSGSDGNFFDCNICLDLSKEPVLTCGHLIYCWPCLYQLQISDAKECPVCKG
EVTSTKVTPIYGRGNHKEIEESLDTKVPMPHARRIESLRNTIQRSPFTIPMEEMIRRI
QNRFDSDSTVPVDFSNREASERVNDRANSILNRLMTSRGVRSEQNQASAAAAAIVAASED
IDLNPNIAPDLEGESNTRFHPILLIRRLQSHRVARISTFTSALSSAERLVDAYFRTHPLG
RNHQEQNHHPVVVDDRDSFSSIAAVINSESQVDTAVEIDSMALSTSSSRRRNENGSRVS
DVDSADSRPPRRRRFT*

>G2701 (46..837)

GTGTTTGTAGTTGAAACTTATCTCTCCCTTTTTTTGTTTTTAGGTATGGAGACTCTGCAT
CCATTCTCTCACCTACCTATCTCTGACCACCGGTTCTGTTTCAAGAGATGGTGAGCTTA
CACAGCTCGAGTAGCGGTAGCTGACTAAAGAAGAGAAACAAGATGTTTCAACGAGCTCTT
GCGATATACGCTGAAGACTCGCCTGATCGCTGGTTTAAAGTTGCTTCCATGATCCCTGGA
AAGACTGTTTTTGATGTTATGAAGCAATATAGTAAGCTTGAAGAAGACGTTTTTCGATATT
GAAGCAGGACGTGTTCCCATTCCTGGTTATCCTGCAGCTTCTTCTCCCTTGGGGTTTGAC
ACGGACATGTGTCTGTAACCGGCCTAGTGGAGCTAGAGGATCTGATCAAGATCGAAAGAAA
GGAGTCCCTTGGACAGAGGAAGAACAACAGGAGATTCTGTTAGGCCCTTCAAGTACGGT
AAAGGAGATTGGAGAAACATATCGAGAAACTTCTGTTGTCAAAGACGCCAACGCAAGTG
GCGAGCCACGCCAAAAGTATTACCAGAGACAGCTCTCCGGAGCCAAGGACAAAACGAGG
CCAAGTATCCATGACATCACAACCGGCAATCTTCTCAATGCCAATCTCAACCGTTCTTTT
TCCGATCATAGAGATATTCTCCCTGATTTAGGGTTTATCGATAAGGATGATACGGAGGAG
GGAGTAATATTTATGGGTGAGAAATCTCTCTTCAAGAAATCTGTTTTCTCCATCACCAACT
TCATTGGAAGCTGCCATTAACCTTCGCCGGAGAAAATGCTTTCAGTGCCGGAGCTTAAGGC
AACATAGAATCCCCAACTCAGCGGC

>G2701 Amino Acid Sequence (domain in AA coordinates: 33-81, 129-183)

METLHPPSHLPISDHRFVVQEMVSLHSSSSGSWTKEENKMFERALAIYAEDSPDRWFKVA

SMIPGKTVFDMVKQYSKLEEDVFDIEAGRVPIPGYPAASSPLGFDTDMCRKRPSGARGSD
QDRKKGVPTWTEEEHRRFLGLLLKYGKGWNRNISRNFVVSKTPTQVASHAQKYYQRLSGA
KDKRRPSIHDITTNLLNANLNRSFSDHRDILPDLGFDIKDDTEEGVIFMGQNLSSENLF
SPSPTSFEAAINFAGENVFSAGA*

>G1819 (1..639)

ATGGAAGAGAACAACGGCAACAACCACTACCTGCCGCAACCATCGTCTTCCCAACTG
CCGCCGCCACCATTTGTATTATCAATCAATGCCGTTGCCGTCATATTCAGTCCGCTGCCG
TACTCACCGCAGATGCGGAATTATTGGATTGCGCAGATGGGAAACGCAACTGATGTAAAG
CATCATGCGTTTCCACTAACCAGGATAAAGAAAATCATGAAGTCCAACCCGGAAGTGAAC
ATGGTCACTGCAGAGGCTCCGGTCTTATATCGAAGGCTGTGAGATGCTCATTCTTGAT
CTCACAATGCGATCGTGGCTTCATACCGTGGAGGGCGGTCCGCCAAACTCTCAAGAGATCC
GATACGCTCAGGATCCGATATCTCCGCCGCAACGACTCGTAGTTTCAAATTTACCTTC
CTTGGCGACGTTGTCCCAAGAGACCCTTCCGTCGTTACCGATGATCCCGTGCTACATCCG
GACGGTGAAGTACTTCTCCGGGAACGGTGATAGGATATCCGGTGTTTGATTGTAATGGT
GTGTACGCGTCACCGCCACAGATGCAGGAGTGGCCGGCGGTGCCTGGTGACGGAGAGGAG
GCAGCTGGGGAATTTGGAGGAAGCAGCGCGGTAATTGA

>G1819 Amino Acid Sequence (domain in AA coordinates: 46-188)
MEENNNGNNHYLPQPSSSQLPPPLYQSMPLPSYSLPLPYSPQMRNYWIAQMGNATDVK
HHAFPLTRIKKIMKSNPEVMNVTAEPVLISKACEMLILDLTMRSLHTVEGGRQTLKRS
DTLTRSDISAATRSFKFTFLGDVVRDPSVVTDDPVLHPDGEVLPPGTIVIGYPVFDNCG
VYASPPQMGEWPAVPGDGEEAAGEIGSSSGN*

>G1227 (372..1451)

TCTTCCGTGTGTTAACAGAAGTCCCCACAATTGTCTGTCTTCGCTGCGAGACAAAACCTGC
CACAGCCAATAATGTTTCTCTGAGGGACCTTGCTTCTGTGACAGACTCGCTCTCTCTCTC
CTCTTCTTGTCTCTGCTCAGCTCTCTCACCAACTCATCTTCAGTCCCTCAAACAAACATCTG
TTCTCATCTTTGTTTCTTTCTTTCTTTCTCATATCTCATTTTCAATTTTCCCAATTTT
TCTTCAACATCTTCATAGCAATTTAAGACCACTATTCCATTATAAAGCTAACTGCTTTAG
AAACTCCTCACATTTATTTCTTCCCATCATTTGTTTTAGAGAGGGAGAAAGAAAAAGAGC
TCAGCTTTCTGATGGAGAGGAGTATTCAAGGACAAAACAAGCTCTGTTGTTTGGACCAA
AAGTGAATGTGAGAAGAAGCCTACAAGTTCAAGAACTGTAGAGGATCATCAAAGCTTTG
CCCTTGAAGAGGAAGAACAACAACCTCTCAACTCCGAGCTTGCTGCAAGACACAACAATAC
CATTTCTACAAATGCTGCAACAAAGTGAAGACCCTTACCCTTTTGTGTCATTCAAAGACC
CAAGCTTTCTAGACTACTATCTCTCCAGACACTTGAAAAGCCTTGGGAACTCGAAAACCT
ACCTCCACATGAAGTTCCAGAGTTTTCATTACCGATCCATTCTGAAACCAACCACTACT
ATCATAATCCATCTTTGGAAGGAGTCAATGAAGCCATCTCAAACCAAGAACTTCCATTCA
ACCCACTAGAGAATGCGCGTTCAAGACGCAAGCGGAAAAACAACAACCTTGGCATCATTGA
TGACAAGAGAAAAGCGAAAGAGAAGAAGAACTAAACCAACAAGAACATAGAAGAGATAG
AGAGTCAAAGAATGACACACATTGCGGTTGAACGAAACCGCAGACGCCAAATGAACGTTT
ATCTGAACCTCACTCCGCTCCATCATTTCCATCTTTCATACATCCAGAGGGGAGACCAAGCGT
CAATAGTAGGAGGAGCAATAGACTTCGTAAAGATCCTAGAGCAACAGTTGCAATCCCTTG
AAGCACAAAAGAGAAGTCAACAGAGTGATGATAACAAAGAGCAAATTCAGAAGATAACA
GTCTCAGGAACATTTGTCGAACAAGTTGCGTGCGAGTAATAAAGAAGAACAAAGTAGCA
AACTCAAATCGAAGCCACAGTGATAGAGAGTCACGTCAACCTAAAAATTCATGTACGA
GGAAACAAGGACAACCTTCTCAGATCAATCATATTGCTGGAGAACTTCGATTCACTGTTT
TTCATCTCAACATCACATCTCCGACCAATACATCTGTCTCTTATTCTTCAACCTCAAGA
TGGAAGATGAATGTAATTTGGGATCAGCGGATGAGATAACGGCGGCGATTCTGTCAGATTT
TCGACAGCTGATTGACTAATCCAAGTAAAAAGTAAAAATAAAAAAGAAACGTTTACTTTG
GTAACCTCGTTTTCATGATTAAATCTTTATTTGGTCGTATGTGATTGGAGTCTTCTCGG
CATGGAACCTGACTTTGGTTTTAGGGTACTAGTCTCTACAGAAGCTGTGGTCCTTCTTTG
GATGC

>G1227 Amino Acid Sequence (domain in AA coordinates: 183-244)
MERSIQGQNKLCCLDQKVNVRSLQVQETVEDHQSFALLEEEQQLSTPSLLQDTTIPFLQ
MLQQSEDPSPLFSFKDPSFLALLSLQTLKPEWELNYLPHEVPEFHSPIHSETNHYHNP
SLEGVNEAISNQELFPNPLENARSRRKRKNNNLASLMTREKRKRRTKPTKNIEEIESQR
MTHIAVERNRRRQMVHLNSLRSIIPSSYIQRGDQASIVGGAIDFVKILEQQQLQSLEAQK
RSQQSDDNKEQIPEDNSLRNIISSNKLRSNKEEQSSKLKIEATVIESHVNLKIQCTRKQG
QLLRSIILLEKLRFVLHLNITSPNTSVSYFNLKMEDECNLGSADEITAAIRQIFDS*

>G2417 (118..1311)

CATACCGGTGGAAGATTCTGCTTTACTACGCTCTCCGCTTCTTCTTCTCCTCGATTTCGAT
TCTCCTCATGGGTTTATCATGAATTTTAGGTTTGTAGTAATTCAGAACTCGAGTGATG
ATCCCGAATGATGATGATGATGCAAATCTATGAAGAATTATCCGTTAAATGATGATGAT
GCAAATCTATGAAGAATTATCCGTTAAATGATGATGATGCAAATCTATGGAGAATTAT
CCGTTAAGGTCAATTCCGACGGAGCTTTCACACACTTGTTTCATTGATACCACCTTCTTTA
CCAAACCCCTTCAGAAGCAGCAGCAGACATGTCCTTCAATTCAGAACTCAATCAAATCATG
GCAAGGCCTTGATATGCTCCCTGCCAATGGTGGAGCTGTTGGTCATAACCCCTTTTTTG
GAACCAGGATTCAACTGCCCCGAGACAACAGATTGGATTCCCTCTCCACTCCCCCATATT
TATTTTCTCTCGGGTTCTCCCAATCTAATAATGGAGGATGGTGTCTATTGATGAGATTCAC
AAACAAAGTGACTTGCCACTTTGGTATGACGACTTGATTACCACTGATGAAGATCCACTC
ATGTCTAGTATCTTGGGCGATCTTCTCCTTGACACTAATTTCAACTCAGCTTCAAAGGTC
CAGCAACCAAGTATGCAATCGCAGATTCAACAACCCCAAGCTGTTCTGCAGCAGCCTTCT
TCTTGTGTGGAATTGCGCCCACTTGATAGGACAGTATCCTCAAACAGCAACAACATAGC
AACAGTAATAATGCAGCAGCAGCAGCTAAGGGACGTATGCGTTGGACGCTGAACCTCAT
GAGGTTTTTGTGACGCTGTTAACCAGCTCGGTGGCAGTAATGAAGCAACTCCTAAAGGT
GTCCTGAAGCATATGAAAGTCGAAGGTTTGACTATTTTTTCATGTCAAAAGTCATTTGCAG
AAATATAGAACAGCTAAATATATACCAGTACCATCAGAAGGTTCCGCGGAGGCAAGGTTG
ACACCGCTTGAGCAAATTACATCTGATGATACGAAACGTGGGATAGATATCACTGAGACT
CTGCGAATTTCAGATGGAACATCAGAAGAACTGCATGAGCAGCTTGAGAGTCTAAGAACA
ATGCAACTTCGGATAGAAGAGCAAGGAAAGGCGCTGTTGATGATGATTGAGAAGCAAAAT
ATGGGTTTCGGCGGACCAGAACAAAGGAGAGAAAAAAGTGCAGAAACGCTGAAAATGGT
TCAGAGGAGTCGGAATCCCCGCGCCAAAGCGTCCGAGAAATGAAGAATGAAGGAAACCT
TTCTTCGGATGGTAGATCATAAACTGTGGTTTTGGTGGAGTTGTAGAGTATGACTTATT
AGGAGTAGAGCTTTCAGTCTTCTTCAGGC

>G2417 Amino Acid Sequence (domain in AA coordinates: 235-285)

MI PNDDDDANSMKNYPLNDDDDANSMKNYPLNDDDDANSMENYPLRSIPTELSHTCSLIPPS
LPNPSEAAADMSFNSELNQIMARPCDMLPANGGAVGHNPFLPEPGFNCPETTDWIPSPPLPH
IYFPGSPNLIMEDGVIDEIHQSDLLPLWYDDLITDEDPMLSSILGDLILLDTNFNSASK
VQQPSMQSQIQPQAVLQPPSSCVELRPLDRVTSSNSNNNSNNNAAAAAKGRMRWTPPEL
HEVFDVAVNQLGGSNEATPKGV LKHKMKEGLTIFHV KSHLQKYRTAKYIPVPSEGSPEAR
LTPLEQITSDDTKRGIDITETLRIQMEHQKLHEQLESRLTMQLRIEEQKALLMMIEKQ
NMGFGGPEQGEKTSAKTPENGSESESPRPKRPRNEE*

>G2116 (104..1117)

TTCACTCCATCATTATCTCCATTGACATTGTTCTCAATTGCGAATAATAATCATAATTA
TTCACACAACCAAGCAATTCATCTCTCAGATTCTCTTAAAAAATGGAGAAATCAGATCC
TCCACCAGTCCCAAAGCCCGCGCCACTATTATCCCCTCCTCCGATCCAATTCTTAATGC
CGATCCGATTCCATCTTCTTCTTCCACCGCCGATCTCGTCCGACGATATGTCCATGTT
CATGTTTCATGATCCCTCTCCTCCGCGCACCACCTTCTTCCGACGACCTTCCCTCCGA
CGACGATCTCTTCTCTTCTTTCATCGATGTCGATAGCCTCACCTCTAATCCCAATCCCTT
TCAAAATCCTTCCCTCTCCTCCAACCTCCGTTTCCGCGCTGCTAATCCTCCTCCTCCTCC
TTCCTCTCGTCTCGCCACCGTCACAGCAATTCCGTTGACGCTGGATGCGCATGTATGC
CGGTGATATCATGGACGCTAAGAAAGCTATGCCTCCTGAAAACTCTCTGAGCTTTGGAA
CATCGATCCCAAACGCGCCAAAAGGATTCTAGCGAATCGACAATCTGCAGCTCGATCCAA
AGAGAGAAAAGCTCGATACATTCAAGAACTTGAGCGCAAAGTTCAATCTCTTCAAACCGA
AGCTACCACTCTCTCTGCTCAGCTTACTCTCTACCAGAGAGACACAAATGGACTAGCAAA
CGAAAACACAGAGCTGAAACTTAGGTTGCAAGCAATGGAACAACAAGCTCAGCTTCGTAA
TGCTTTTAAACGAAGGTTTGAGGAAAGAGTTGAAAGGATGAAGATGGAGACAGGAGAAAT
CTCTGGTAATTCAAGATTCTGTTTGATATGGGAATGCAGCAGATTCAATCTCTTCTCAAC
TTTCATGGCTATTCCACCATATCATGGCTCAATGAACCTCCATGATATGCAGATGCATT
TAGTTTCAATCCTATGGAGATGTCCAATTCTCAAAGCGTGTCCGACTTTCTACAGAACGG
CCGAATGCAAGGGCTGGAGATTAGTAGCAATAGCTCAAGCTTAGTCAAATCTGAAGGACC
TTCTCTCTCTGCTAGTGAGAGTAGCTCTGCCTATTGACGACAAGATTATGATGAGGCTCA
TTTTTCTG

>G2116 Amino Acid Sequence (conserved domain in AA coordinates: 150-210)

MEKSDPPVPKPGATIIIPSSDPIPNADPIPSSSFHRRSRSDMSMFMDPLSSAAPPSS
DDLPSDDDLFSSFIDVDSLTSNPNPFQNPSSLSSNSVSGAANPPPPSSRPRHRHSNSVDA

GCAMYAGDIMDAKKAMPPEKLSELWNIDPKRAKRILANRQSAARSKERKARYIQELERKV
QSLQTEATTLAQLTLYQRDTNGLANENTELKLRLQAMEQQAQLRNALNEALRKEVERMK
METGEISGNSDSFDMGMQIQYSSSTFMAIPYPYHGSMNLHDMQMHSFNPMEMSNSQSVS
DFLQNGRMQGLEISSNSSSLVKSEGPSLSASESSAY*

>G647 (1..948)

ATGATGATCGGCGAAAATAAAAACCGGCCACATCCAACGATCCATATCCCTCAATGGGAT
CAAATCAACGATCCAACCGGCCACAATCTCTTACCATTCTCTTCCGTCAACCTTAACAGC
GTTAACGACTACCCACACTCTCCGTACCGTATCTCGACTCCTTCGCTTCTCTCTCCGT
TACCTCCCGTCAAACGAGTTAAACAAACGATTGAGACTCATCAAGTGGCGACGAGTCATCA
CCACTCACCGACTATTCTCTCCGACGAGTTTCGCATCTACGAGTTCAAAATCCGCGCA
TGCCTCGAGGTCGATCTCATGATTGGACGGAGTGTCCGTTTCGCACATCCCGGAGAAAAA
GCTCGACGACGTGATCCGAGAAAGTTTTCATTACTCCGGCACCCTTGTCTGAGTTTCTGT
AAAGGAAGTTGTAGAAGAGGTGATTCTGTGAGTTCTCTCATGGAGTTTTCGAGTGTG
CTCCATCCTTCTCGTTACCGTACTCAGCCGTGTAAAGACGGAAGTAGCTGCCGAGAAAGA
ATCTGTTTCTTCGTCATACGACGGAGCAGTTACGTGTATTACCTTGTTCGTTAGATCCA
GATCTTGGATTCTTCTCAGGATTAGCTACTTCTCCGACTTCGATTCTTGTTCCTCTCCG
TTTTCAACCACCGTCGGAATCTCCGCCGCTTCTCCGAGTACCGTGAACCTTATTGCGTCG
ATGAGGAAAATGCAATTGAACGGAGGTGGTTGTTTCGTGGAGTTCTCCGATGAGATCTGCA
GTTAGGTTACCTTTTTCTGTCGTCTCTGCGTCCGATTACAGCGGCAACGTGGCCGAGGATA
AGAGAGTTTGAATCGAAGAAGCTCCGGCGATGGAATTTGTGGAATCTGGGAAAGAGCTG
AGAGCGGAGATGTATGCAAGACTCAGTAGAGAGAACTCACTCGGTTGA

>G647 Amino Acid Sequence (domain in aa coordinates: 77-192)
MMIGENKNRPHPTIHIPQWDQINDPTATISSPFSSVNLNSVNDYPHSPSPYLDSPFASLFR
YLPNSNELTNDSDSSSGDESSPLTDSFSSDEFRIYEFKIRRCARGRSHDWTECPFAHPGEK
ARRRDPKRFHYSGTACPEFRKSGSCRRGDSCEFSHGVEFCWLHPSRYRTQPCKDGTSCRRR
ICFFAHTTEQLRLVLPCLSDPDLGFFSGLATSPTSILVSPSFSPSESPPLSPSTGELIAS
MRKMQLNGGCSWSSPMRSVRLPFSSSLRPIQAATWPRIREFEIEEAPAMEFVESGKEL
RAEMYARLSRENSLG*

>G974 (377..1162)

AAAAAAAAAGTTGATATACTTTCTGGTTTTCTCCTTAACTTTTATTCTTTACAAATCCAT
CCCCCTTAGATCTGTTTATTTCCCGCTACTTTGATTCAATTTCTGTTAGTAATCTGTCTTT
CGTATAGAAGAAAATGATTCTTGGTTTGTATTTTCTTAAAGAGATCAATCTTTTTTTA
TTTTTGATCTTCTGTGTTTTTTTCTTTGTAGAAATTAATCGTTTGTGAGGGTATTTTT
TTAATTCCTTCTCTCAGAAATCTACACAGAGGTTTTTTATTTTATAAACCTCTTTTTCTG
ATTTTCTTGAAAAACAAAAATCCTGTTCTTTACTTTTTTTACAAGAACAAGGAAAAAAA
TTTCTTTTTATTAGAAATGACAACTTCTATGGATTTTTTACAGTAACAAAACGTTTCAACA
ATCTGATCCATTCCGTGGTGAATTAATGGAAGCGCTTTTACCTTTTATCAAAAGCCCTTC
CAACGATTCATCCGCGTTTTCGCTTCTCTACCCGCTCCAATTTATACAGGGTCGGATCT
CCACTCATTTTCTCACCATCTTAGTCCTAAACCGGTCTCAATGAAACAAACCGTACTTC
CGCGGCTAAACCGACGAAGCTATACAGAGGAGTGAGACAACGTCACTGGGGAAAATGGGT
GGAGGAAGCTGCTTTAGCTTATGACAAGCGCGGTATAAGCTCCGAGGAGATTTTTCGCG
GCTTAATTTCCCTGATCTCCGTCTAAGCAGAGTATCAACCTCTTCAATCATCAGTCGA
CGCTAAGCTTGAAGCTATTTGTCAAACTTAGCTGAGACGACGAGAAACAGGTGAGATC
AACGAAGAAGTCTTCTCTCGGAAACGTTTCAACCGTCGAGTGAAACTACCGGAGGA
GGACTACTCTAGCGCCGGATCTTCGCCGCTGTTAACGGAGAGTTATGGATCTGGTGGATC
TTCTTCGCCGTTGTCGGAGCTGACGTTTGGTGATACGGAGGAGGAGATTACGCCCGGTG
GAACGAGAACGCGTGGGAGAAGTATCCGTCGTACGAGATCGATTGGGATTGATTCTTCA
GTGTTCCAGTCTTGTAATTAGATGTTGCCATAGGGGTATTTTAGGGACTTTAGAGCTCT
CTGCGATGGAGTTTTTGGTCATTGCAGAGATTTTATTATTATTAAAGGGGGTTTGTATGT
TAATATCAAATAAGTTTATCTACTTTGATGTTAATTAGTGTAAATCTCTGCGTCGGTCCA
AGCTGTTTTTTTTTGGCATGCTTCGACCGTGTGAGATTTCTTATGTAATTTTTGTAGTTC
CTTGATTTTCTTAGTTCAAGTTAAATTGGCACAAAAAATAAAAAAAAAA

>G974 Amino Acid Sequence (domain in AA coordinates: 81-140)
MTTSMDFYSNKTFFQSDPFGGELMEALLPFIKSPSNDSSAFAPISYGSDDLHSFSH
HLSPKPVSMKQTGTSAAKPTKLYRGVQRHWGKWVAEIRLPRNRTRLWLGTFTDAEEAAL
AYDKAAYKLGRDFARLNFPLRHNDEYQPLQSSVDKLEAICQNLAEATTQKQVRSTTKSS

SRKRSSTVAVKLPEEDYSSAGSSPLLTESYSGSGSSSPLSELTFGDTEEBIQQPPWNENAL
EKYPSYEIDWDSILQCSSLVN*

>G1419 (27..692)

GAAGACTCCAACATAATTCATCATCTATGGCTTCTTCACATCAACAACAGCAAGAACAAG
ACCAGTCAGCTTTAGATCTCATAACCCAACACCTTCTTACTGATTTCCCTTCCTTAGACA
CCTTTGCCTCCACCATCCACCCTGCACCACCTCAACTCTAAGCCAACGCAAACCACCTC
TTGCCACTATAGCAGTTCTTACTACTGCACCGGTGGTTCAAGAGAATGATCAAAGGCATT
ACAGAGGCGTCAGGAGAAGACCATGGGGTAAGTATGCGGCTGAGATCAGAGACCCAAACA
AGAAAGGTGTTCTGTCTGGTTAGGCACCTTTTGACACAGCCATGGAAGCTGCAAGAGGTT
ATGACAAGGCAGCTTTTAAACTACGAGGAAGCAAAGCTATTCTTAACTTCCCACTTGAAG
CAGGAAAGCATGAGGACTTGGGAGACAACAAGAAGACTATTTCTTTAAAGCAAAGAGGA
AGAGACAGGTGACGGAGGATGAAAGCCAGCTGATCAGCCGTAAAGCTGTTAAGAGGGAAG
AAGCTCAGGTTACGGCTGATGCTTGTCATTACGCCATCAAGTTGGAAGGGGTTTTGGG
ACGGAGCAGACAGTAAAGACATGGGAATATTTCCGTGCCTCTGTTATCTCCTTGTCAT
CTCTTGACACTCTCAACTCGTAGTTACTTAAGCTTCAGAGGGTCAAAGTGGAAAAATC
AACATTGGATTGTTTTCAAAGCTTCTAGATTAGCTGATTGTAAAAAATGTTTTACTATA
TTCATTCACTTCTTAAATGCAATTCCTTCTACCTTCC

>G1419 Amino Acid Sequence (domain in AA coordinates: 69-137)

MASSHQQQQEQDQSALDLITQHLLTDFPSLDTFASTIHHCTTSTLSQRKPPLATIAVPTT
APVVQENDQRHYRGVRRRPWGKYAAEIRDPNKKGVRVWLGTDFDAMEAARGYDKAAFRLR
GSKAILNFPLEAGKHEDLDGNKKTISLKAKRKRQVTEDESQILSRKAVKREEAQVQADAC
PLTPSSWKGFWDGADSKDMGIFSVPLLSPCPSLGHSLQLVVT*

>G1634 (22..855)

TTATCTCGTAGCCTTTAAACGATGGAGACTCTGCATCCACTACTCTCGCACGTGCCAACT
TCTGACCACCGGTTTGTAGTTCAAGAGATGATGTGCTTGCAAAGCTCGAGCTGGACTAAA
GAAGAGAACAAGAAGTTTGAGCGAGCTCTTGCTGTCTACGCTGATGACACGCCTGATCGC
TGGTTCAAAGTTGCTGTATGATCCCTGGAAAGACCATATCAGATGTCATGAGGCAATAC
TCTAAGCTTGAAGAAGACCTCTTCGATATCGAAGCAGGACTTGTCCCGATCCCGGGTTAC
CGTTCAGTTACTCTTGTGGATTGATCAGGTTGTGAGTCCACGTGACTTTGATGCGTAT
CGTAAACTTCTAATGGAGCCAGAGGATTGATCAAGACCGTAGGAAAGGAGTTCCATGG
ACGGAGGAAGAACAACAGGAGATTCTTGTAGGGCTTCTCAAGTATGGGAAAGGAGATTGG
AGAAAACATATCGAGGAACCTTGTGGGATCAAAAAACCAACTCAGGTTGCAAGTCATGCC
CAAAAGTACTACCAAAGACAGCTTTCCGGTGCGAAAGACAAACGACGGCCTAGCATTCAC
GACATCACACCGTCAATCTTCTCAATGCCAATCTTAGCCGTCCATCGTCTGATCACGGT
TGCTTAGTCTCAAAACAGGCCGAGCCGAAACTAGGGTTACCGACAGGGATAATGCAGAG
GAGGGAGTTATGTTTCTTGGTCAGAATCTATCCTCGGTCTTCTTCTTCTACGATCCTGCC
ATTAAGTTTTCCGGAGCAAATGTTTACGGTGAAGGAGGTTACTGTATCTCACAAGATCTT
GAAACGAGAAATGAGAATTTTGAATTTTAACTATTGCAACGAAACCATAATTGC

>G1634 Amino Acid Sequence (domain in AA coordinates: 129-180)

METLHPLLSHVPTSDHRFVVQEMMCLQSSSWTKEENKKFERALAVYADDTFDRWFKVAM
IPGKTISDVMRQYSKLEEDLFDIEAGLVPIPGYRSVTPCGFDQVVSPRDFDAYRKLPGA
RGFDQDRRKGVPTWTEEBHRRFLLGLLKYKGDWRNISRNFVGSKTPTQVASHAQKYYQRQ
LSGAKDKRRPSIHDITTVNLLNANLSRPSSDHGCLVSKQAEPKLGFTDRDNAEEGVMFLG
QNLSSVFSSYDPAIKFSGANVYEGGGYCISQDLETRK*

>G1637 (1..954)

ATGGTGAAGGAGACGGTGACGGTGGCGAAAAACGTGCTCACACTGTGGCCATAATGGCCAT
AACGCACGGACTTGTCTCAACGGCGTTAATAAGGCAAGTGTAAACTGTTTCGGCGTTAAT
ATATCGTCTGATCCGATTAGGCCCGCTGAGGTAACGGCGTTAAGGAAGAGTCTTAGTTG
GGAAACCTTGATGCTCTTCTCGCTAACGATGAAAGTAACGGTAGCGGTGATCCTATCGCC
GCCGTTGATGATACCGGTTATCATTCCGATGGTCAGATTCAATCCAAGAAGGGTAAACT
GCTCATGAGAAGAAAAAGGGGAAGCCATGGACGGAAGAAGAACATCGTAATTTCTTAATC
GGTTTAAACAACTCGGAAAAGGAGATTGGAGAGGCATTGCAAAGAGTTTCGTGTCGACA
AGAACACCAACACAAGTCGCAAGTCATGCTCAGAAATATTTTATTAGGTAAACGTTAAC
GACAAGAGAAAAAGACGTGCTAGTCTCTTACATCTCTCAGAGATCAGAAGGAGAAA
GAGAGGAAGCTCTAAGATGCTTCAACAAAGACTCCACCTAAACAACCAATAACCGGAATT
CAACAACCGGTAGTACAGGTCATACCTCAAAACCGAGATTTCGAACAGGTTTCAGAATTTA
TCAATGGAGTATATGCCAATCTACCAACCCATACCACTTACTACAACCTTTCCACCTATT

ATGTACCATCCAAATTATCCAATGTACTATGCCAACCTCAAGTACCGGTTAGGTTTGTTCATCCTTCTGGTATACCTGTTCCAAGACATATACCGATTGGTTTGCCTCTGTCTCAACCGAGTGAAGCTTCTAATATGACAAATAAAGACGGTTTGGATCTTCATATCGGTTTGCCTCCA

CAAGCTACTGGAGCTTCTGACTTGACTGGTCATGGCGTTATTCATGTGAAATGA
>G1637 Amino Acid Sequence (domain in AA coordinates: 109-173)
MVKETVTVAKTCSHCGHNGHNARTCLNGVNKASVKLFGVNISSDPIRPEVTALRKSLSL
GNLDALLANDESNGSGDPIAAVDDTGYHSDGQIHSKKGKTAHEKKKGKPWTEEEHRNFLI
GLNKLKGKDWRIAKSFVSTRPTQVASHAQKYFIRLNVNDKRKRRLASLFDISLEDQKEK
ERNSQDASTKTPPKQIPITGIQPPVVGHTQTEISNRFQNLMEYMPIYQPIPPYYNFPPI
MYHPNYPMYANPQVPVRFVHPSGIPVPRHIPGLPLSQPSEASNMTNKGDLDLHIGLPP
QATGASDLTGHGVIHVK*

>G1818 (601..1161)
TAACAAATCAAATAATTAGAGAAATAACCAAAATTTAACTTTTAGAGGGACTACAGGATT
TGTACTTTGTACATTTCATATATTATTGTTATATATCGTTTCATACATTAATTTGAACCAA
TGTAATAATAGTAAATTTCAATTTAACATCATGAGCAAATTCCTATTAAATTTCTCTTAA
AATTTTGAGCAAATTATGCTTTCACATTTAATTTGAAAACATCATTTTAAACAAGATA
TTCAAACCTAAGTTTGTACAGCAAATTTTAACTTTCAATTTTATAGAGAAAAAGGTAT
TTTTTTTTTTGTTTCATTTTATAAGACTATTATTGTTATATAATATACACTTTAAGTA
AAAACAAATCTCTTTCTTTTCTTCTTATAATACCAACCACAAGTCTGTCAGTCACACA
CATACAGTTAATAACATTTAAATATTCTTAACAACTACTAAATAGGTTGAGATTTCATATA
TGTAAGAGATCACTTCTTAATCTTATCCTACCATATCTTATATACGCTTAATTTTCCTT
TATATATGCAAACCTCCACATAAAAAATATCTCAAACCCAAACACTTCAAACAAAAA
ATGGAGAACAAACAACCAACCAACCAACGAGCCAGGAAAGATAACGAGCAACTAAAGAGT
TTCTGGTCAAAGGGGATGGAAGGTGACTTGAATGTCAAGAATCAGAGTTCCCCATCTCT
CGTATCAAGAGGATAATGAAGTTTGATCCGGATGTGAGTATGATCGCTGCTGAGGCTCCA
AATCTCTTATCTAAGGCTTGTGAAATGTTTGTGATGGACCTCACGATGCGTTCATGGCTC
CATGCTCAAGAGAGCAACCGACTCACGATACGGAATCTGATGTTGATGCCGTAGTGTCT
CAAACCGTCATCTTTGATTTCTTGGCTGATGATGTCCCTAAGGACGAGGGAGAGCCCGTT
GTCCGCGCTGCTGATCCTGTGGACGATGTTGCTGATCATGTGGCTGTGCCAGATCTTAAC
AATGAAGAACTGCCGCCGGGAACGGTGATAGGAACCTCCGGTTTGTACGGTTTAGGAATA
CACGCGCCACACCCGCAGATGCCTGGAGCTTGGACCGAGGAGGATGCGACTGGGGCAAAT
GGAGGAAACCGGTGGGAATTAATATTGGATTGGGTTTGTAAACCGCTGTTGTGAGAATT
GAATTTCTTTTGGATTCTGCTTATGTTTTCATGTTTATGTTTTTAGTTGTTGAATGTA
TTTCTGTTGTTTTGTCCAAAAAAGAAATGTATTCTGTTGTTGCTTTCAAATGA
ATCTAATGGTTTATGAATATTGGCTTTAGATTAAATTTATGCATACAAAACACAAGGATT
ACGGATAAAAAAGTCTCAGTTTACCCATGGAAACATAATCTTCTAGTGATTCTTTATGA
GAGTAGAAAAGAAATCATATATTATAATCTATTTTCATAAGAGATAGGGTACTGTAAACAAG
GATGTTTATTCCGGCTATTTCTTTTTTTTTTAATCACTTTTACTTGTCAAGACTCTTTTGT
GTTTGCAGCTTTTGTGATTACATTCTAGAGGCAACAAGATCCAGAGATCTAGCAAAA
AAAACCTATTGTGAACCTGAATCTATTTAAAAATTTCCAACTCATTTTTCGTTCTTA
TTCTTTGTTTTCACCGGAATTTGGCGCACAAACGATTATTGTAATTTTGTCTTTCAAG

>G1818 Amino Acid Sequence (domain in AA coordinates: 36-113)
MENNNNNHQPPKDNEQLKSFWSKMEGDLNVKNHEFPISRIKRIMKFDPDVSMIAAEAP
NLLSKACEMFVMDLTMRSLHAQESNRLTIRKSDVDAVVSQTVIFDFLRDDVPKDEGEVP
VAAADPVDDVADHVAVPDLNNEELPPGTVIGTPVCYGLIHAPHQPMPGAWTEEDATGAN
GGNGGN*

>G1820 (1..609)
ATGGCTGAGAACAAACAACGCGGACAAACATGAACAACGACAACCACCAGCAACCA
CCGTCGTACTCGCAGCTGCCGCGATGGCATCATCAACCCCTCAGTTACGTAATTACTGG
ATTGAGCAGATGGAACCGTCTCGGATTTCAAAAACCGTCAGCTTCCATTGGCTCGAATT
AAGAAGATCATGAAGGCTGATCCAGATGTGCACATGGTCTCCGCAGAGGCTCCGATCATC
TTCCGAAAGGCTTGCGAAATGTTTCATCGTTGATCTCACGATGCGGTCGTGGCTCAAAGCC
GAGGAGAACAAACGCCACACGCTTCAGAAATCGGATATCTCAACGCAGTGGCTAGCTCT
TTCACCTACGATTTCTTCTTGATGTTGTCCCTAAGGACGAGTCTATCGCCACCGCTGAT
CCTGGCTTTGTGGCTATGCCACATCCTGACGGTGGAGGAGTACCGCAATATTATTATCCA
CCGGGAGTGGTGATGGGAACCTATGGTTGGTAGTGGAATGTACGCGCCATCGCAGGCG
TGGCCAGCAGCGCTGGTGACGGGGAGGATGATGCTGAGGATAATGGAGGAAACGCGCGC

GGAAATTGA

>G1820 Amino Acid Sequence (domain in AA coordinates: 70-133)
MAENNNNNNGDNMNNNDNHQQPPSYSQLPPMASSNPQLRNYWIEQMETVSDFKNRQLPLARI
KKIMKADPDVHMVMSAEAPIIFAKACEMFIVDLTMRSLWKABENKRHTLQKSDISNAVASS
FTYDFLLDVPVKDESIA2DPGFGFVAMPHPDGGGVPPQYYYPPGVVMGTPMVGSGMYAPSQA
WPAAAGDGEDDAEDNGGNGGGN*

>G1903 (1..1200)

ATGTCTAAATCTAGAGATACGGAGATAAAGTTGTTTTGGGAGGACAATCACATCTCTTTTAT
GATGTGAATTGTTATGATCCGTCGTCGTTGTCCCCTGTTCCAGATGTTTCTTCTGATCCA
AGCAAGGAGGATTCGTCTTCTTCTTCATCTTCTTGTTCTCCAATATTGGACCAATCAGG
GTTCCGGTTAAAAAAGTGAGCAAGAGAGTAACAAATTCAAAGATCCATATATATTATCC
GATCTAAACGAACCACAAAAGCAGTATCTGAGATTTTCATCACCAAGAAGTTCCAAGAAC
AACTGTGATCAACAGAGCGAGATCACAAACAACACTACCACAAGTACTACATCAGGAGAG
AAATCAACGGCTTCTCAAGAAACCGGACAAGCTTATTCATGTCTTAGATGTGAAAGCGCA
AACACAAATTTCTGTTATTACAACTACAACCTGAACCCAGCCAGTTACTTCTGCGAG
AACTGTCAGAGGTATTGGACAGCTGTGGATCATGAGGAACGTTCTCTGTGGCTCAGGT
CGTCGCAAGAACAAAGGATGGCCTTCTTCAAACATTACTTGCAGTCACTTCTGAGGAT
TGTGATAATAATAACTCGGGGACGATCCTTAGTTTCGGTTCTTCGGAGTCTTCGGTTACA
GAGACTGGTAAGCATCAGTCAGGTGATACAGCAAAGATAAGTGCTGATTACAGTTTCTCAA
GAAAATAAAAGCTACCAAGGGTTTCTTCTCCTCGCAAGTAATGTTACCTAATAATTCTTCT
CCTTGGCCTTACCAATGGAGTCCAACGGGTCTTAACGCTAGTTTCTACCCTGTCCCCTTC
TACTGGGGATGCACGGTTCCGATATACCCTACCTCAGAGACTTCATCATGTTTtaggAAAA
CGGTCAAGAGATCAAACTGAAGGAAGAATCAATGATACTAATAACAACAATAACTACTACA
AGAGCAAGATTGGTCTCAGAATCTCTTAGAATGAATATCGAAGCTAGTAAGAGCGCTGTG
TGGTCTAAGTTACCGACAAAACCCGAGAAAAAACCGCAAGGATTCAGTTTGTTCATGGGA
TTTGACACAAAGGGAAACAGCAACAGAAGTAGCTTGTCTCCGAACTTCTCAGCTCTA
CAAGCAAACCTGCAGCGATGTCTAGAGCTATGAACCTTCAGGGAGAGCTGCAACATAA

>G1903 Amino Acid Sequence (domain in AA coordinates: 134-180)

MSKSRDTEIKLFGRTTISLLDVNCYDPSSLSPVHDVSSDPKEDSSSSSSSSCSPITGP
VPVKKEQESNKFDPYILSDLNEPPKAVSEISSPRSSKNCDQOSEITTTTTTSTTSGE
KSTALKPKDKLIPCRPCEASANTKFCYYMNYNVNQPRYFCRNCQRYWTAGGSMRNVPGSG
RRKNKGWPPSNHYLQVLTSEDCDNNNSGTILSFGSSESSVTETGKHQSGDTAKISADSVSQ
ENKSYQGFLPPQVLMPLNNSSPWPYQWSTGPNASFYPVVFYWGCTVPIYPTSETSSCLGK
RSRDQTEGRINDTNTTITTTTRARLVESESLRMIENASKAVWSKLPTKPEKKTQGFSLFNG
FDTKGNSNRSSLVSETSHSLQANPAAMSRAMNFRESMQQ*

>G371 (1..582)

ATGGAGATTGAGAAGGATGAGGACGACACAACATTGGTTGATTCTGGAGGAGACTTCGAC
TGCAACATATGTTTGGATCAGGTTGAGACCCGGTCGTGACTTTATGTGGCCACCTGTTTT
GTTTGGCCCTGCATTCACAAGTGGACTTATGCGTCCAACAATTCAAGACAACGAGTCGAT
CAATACGATCATAAGAGGGAACCAACAAAATGTCCGGTATGCAAACTGATGTCTCCGAG
CTTACGCTTGTCCCGATCTACGGACGAGGACAGAAAGCTCCCCAGTCCGGTTCAATGTA
CCGACGACAGCAACTGGTCCGGTTTATGACTTAAAGAGGAGTTGGTCAACGTTTAGGAGAA
GGGGAGAGTCAACGTTACATGTTATAGAATGCCTGATCCGGTGATGGGTGTGGTATGCGAA
ATGGTATACCGGAGACTATTTGGAGAGTCTTCGAGCAACTGCGACCTTACC CGGATATG
AATGTCCGGTCTAGGCGACGGGCAATGCAGGCTGAGGAGTCAATTAAGCAGAGTCTACTTG
TTCTACTTTGCTTCATGTTTATGTGTCTATTTCTCTTCTAA

>G371 Amino Acid Sequence (domain in aa coordinates: 21-74)

MEIEKDEDDTTLVDSGGDFDCNICLDQVRDPVVTLCGHLFCWPCIIHKWTYASNNRQRVD
QYDHKREPPKCPVCKSDVSEATLVPIYGRGQKAPQSGSNVPSRPTGPVYDLRGVQRLGE
BESQRYMYRMPDPVMGVVCEMVYRRLFGESSNMAPYRDMNVRSRRRAMQAEESLSRVYL
FLLCFMFMCLFLF*

>G597 (255..1310)

AAAATTCTCCTGTAAAATTTAATATTATAAAAGTGGTTTCTTTTCAATTATGTTTATAT
 AATTTTCATCTTTAATCTTAAATCTGGTAACCTTAATGCGCGATCCGCTTTTCTAAAGT
 TTGTGAGAGAGAAGAGATCTAAAAAAATCCACAATTTGTTCAAATCTTGGAGTTAAAT
 CTGAATTTTAGGCCTTGTGCTTAGATTATGGCTTAAAGTTTCAAATTTTCATTTGA
 ATGTGAGAAGAAAATGTCAGGATCTGAGACGGGTTTAAATGCGCGCAGCAGAGAAATCAA

TGCAATTTACAATGGCTCTCCACCAGCAGCAGCAACACAGTCAAGCTCAACCTCAGCAGT
CTCAGAACAGGCCATTGTCTATTGGGTGGAGACGACGGAAGTCTCTTTACAAGCAGCCGA
TGAGATCAGTATCACCACCGCAGCAGTACCAACCCAAGTCTAGCTGGTGAGAATTCTGTCT
TGAACATGAACCTTGCCCGGAGGTGAGTCTGGAGGCATGACTGGAAGTGGAGTGGAGCCAG
TGAAAAAGAGGAGAGGTAGACCGAGGAAATATGGGCCTGATAGTGGTGAAATGTCACTTG
GTTTGAATCCTGGAGCTCCTTCTTTCACTGTCAGCCAACCTAGTAGCGGCGCGATGGAG
GAGAGAAGAAGAGAGGAAGACCTCTGGTTCTTTCTAGCAAAAGGCTCAAGCTTCAAGCTT
TAGGCTCGACTGGAATCGGATTACGCCTCATGTACTTACCGTGCTGGCTGGAGAGGATG
TATCATCCAAGATAATGGCGTTAACTCATAATGGACCCCGTGCTGTGTGTGTCTTGTCTG
CAAATGGAGCCATCTCCAATGTGACTCTCCGCCAGTCTGCCACATCCGGTGGAACTGTTA
CATATGAGGGGAGATTTGAGATTCTGTCTTTATCGGGATCTTTCCATTTGTGAGAGAACA
ATGGTCAAAGAAGCAGGACGGGAGGTCTAAGCGTGTCTATCAAGTCCGGATGGTAATG
TCCTCGGTGGCAGTGTAGCTGGTCTTCTTATAGCAGCATCACCTGTTTCAAGATTGTGTG
GGAGTTTCTTACCAGACGGAGAAAAAGAACCAAAACAGCATGTGGGACAAATGGGACTGT
CGTCACCCGTATTACCGCGTGTGGCCCCAACGCAGGTGCTGATGACTCCAAGTAGCCAC
AATCTCGAGGCACAATGAGTGAGTCATCTTGTGGAGGAGACATGGAAGCCCTATTTCATC
AGAGCACTGGAGGACCTTACAATAACACCATTAACATGCCCTGGAAGTAGCCAAGTGATC
TGTGTGCGCTTAAAACCAACAACCTCCCGTTATTAGAGTGATTTATTTCTACATTGGTT
TAGACTTTCTAGTTCTGATGGTTATTTCTACAGTTGGTTTAGACTTTCTAGTTCTGTTCA
GACAAAAGGAGTTTGATAAATTGACCGACCTATTTTGTGTGTTGAGGTACTTTCAGAAC
CATAGGTGTTTCAGAAATTAGAATGTTCTGTTTAAAAAA

>G597 Amino Acid Sequence (domain in AA coordinates: 97-104,137-144)

MSGSETGLMAATRESMQFTMALHQQQHSQAQPPQSQNRPLSFGDDGTALYKQPMRSVS
PPQQYQPNASGENSVLNMNLPGGESGGMGTGTGSEPVKKRRGRPRKYGPDSGEMSLGLNPG
APSFVTSQPSGGDGEKKRRPPGSSSKRLKLQALGSTGIGFTPHVLTVLAGEDVSSKI
MALTHNGPRAVCLVLSANGAISNVTLRQSATSGGTVTYEGRFEILSLSGSFHLENNQORS
RTGGLSVSLSSPDGNVLGGSVAGLLIAASPVQIVVGSFLPDGEKEPKQHVGMGLSSPVL
PRVAPTQVLMTPSSPQSRGTMSESSCGGGHGSPIHQSTGGPYNNTINMPWK*

>G1009 (28..1704)

AAAAAAAAAAAAAACCATTCCCAAAGATGAAGAACAATAACAACAAATCTTCTTCTTCT
TCTAGCTATGATTCTTCTTTGTCTCCTTCTTCTTCTCATCTCCTCCCAACAGAACTGGCTC
TCTTTCTCTCTCTCCAACAATAACAACAACCTTCAATTCTTCTCAAACCTAATCTCACT
TCCTCCACATCAGATCATCATCTCACCCTTCTCACCCTCTCTCTCTTTCAAGCTTTC
TCCACTTCTCCAGTCCGAACGGCAAGATGGGTACCCGGGAGTTTCACCCAAGATGCCACG
GCGGTTCTTTCCGTATACCCCGGCGGTCTTAACTTGAGAACTTCTCGGCGGAGGAGCC
TCAACGACGACAACAAGACCAATGCAACAAGTGCAATCTTGGCGGCGTTGTCTTCTCT
TCCGACCTACAGCCACCGCTTCATCTCCGTCCGCGCGGAGATCTACGACTCTGAGCTC
AAGTCAATAGCCGCTAGCTTCTTAGGAACTACTCCGGTGGACACTCGTCCGAGGTCTCT
AGCGTACATAAACAACAACCGAATCTCTAGCTGTCTCAGAGGCTTCGCCTACTCCGAAG
AAGAACGTAGAGAGTTTGGACAACGTACCTCGATTTATAGAGGAGTCACAAGACATAGA
TGGACTGGAAGATACGAAGCTCATCTATGGGATAATAGTTGCCGAAGAGAAGGCCAAAGC
AGAAAAGGAAGACAAGTTTATTTAGGTGGTTATGATAAGGAAGATAAAGCAGCTAGAGCT
TACGACCTTGACGCTCTTAAGTATTGGGGTCTTACAACACTACGACTAATTTCCCGATATCA
AATTACGAATCTGAACCTGAAGAAATGAAACACATGACTCGACAAGAGTTCGTTGCTTCT
TTAAGACGGAAGAGCAGTGGATTCTCTAGGGGTGCCTCCATGTACAGAGGCGTCACTAGA
CATCATCAGCATGGTCGATGGCAGGCACGAATTGGAAGAGTTGCAGGCAACAAGACCTT
TATCTTGGCACATTTAGCACTCAAGAGGAAGCTGCAGAAGCTTATGATATAGCAGCGATC
AAATTCCGCGGTCTAATGCAGTCACCAATTTCGACATCAGTCGATATGATGTCAAATCA
ATTGCTAGCTGTAATCTCCCTGTGGGTGGACTAATGCCTAAACCTTCTCCAGCAACCGCA
CGCGCTGACAAAACCGTTGATCTTCTCCATCCGACTCTCCATCTCTAACCACACCGTCC
CTCAGCTTCAATGTGGCAACACCGGTCAATGACCATGGAGGAACCTTTTACCACACTGGT
ATACCAATCAAACAGACCCGCTGATCATTATTGGTCCAACATCTTTGGATTCCAGGCA
AACCCGAAAGCAGAAATGCGACCATTAGCAAACCTTGGGTGCGATCTTCATAACCTTCT
CCTGGTTATGCTATAATGCCGTAATGCAGGAAGGTGAAAACAACCTTGGTGGTAGTTTT
GTTGGGTCTGATGGGTATAACAATCATTCCGCTGCATCGAACCCGGTCTCAGCAATTCCG
CTGTCTCGACAACATAATGAGTAACGGTAACGAAGGGTATGGTGGAAACATAAACTGG
ATTAATAACAACATTTCAAGTTCTTACCAAACCTGCAAAATCAAATCTCTCTGTTTGCAC

ACACCGGTTTTTGGGTTGGAATGAGTATTCACATCTTAGTGAGAACTAAAATAAATATGT
 AGGAAAAAATAAGGCTCTGTTTGAAGAAATCAGATATTTTCTTCTTAGATTATTTAAGT
 AGTTTAAAAAATATTTTTTAAAGTGTTCACCTTTACGTTTGTCTGCTGACCACGAATT
 TTGCTGGATCTGACAGTACTAACTCTTTGTTTAATGACCTTATGGGTTCCTTTTTACTT
 TCCAGAACTTTTATTTACTTTTTTCTTCATTTTTTCTTCATTTTTTTTGTGTGGGACAAT
 ATGAATGATTGAAGATGGAACTGCTTGCATGTGAATAAACGAAATCAAACNATCTTCG
 GTAACCTAAAAA

>G1009 Amino Acid Sequence (domain in aa coordinates: 201-277, 303-371)

MKNNNNKSSSSSSSYDSSLSPSSSSSSSHQNWLSFSLNNNNNFSSSNPNLTSSTSDHHHP
 HPSHLSLFQAFSTSPVERQDGSFVSPSDATAVLSVYPGGPKLENFLGGGASTTTTRPMQ
 QVQSLGGVVFSSDLQPPHPPSAABEYDSELKSIASFLGNYSGGHSSEVSSVHKQQPNP
 LAVSEASPTPKKNVESFGQRTSIYRGVTRHRWTGRYEAHLWDNSCRREQSRKGRQVYL
 GYDKEDKAARAYDLAALKYWGPITTTNFPISNYESELEEMKHMTRQEFVASLRKSSGFS
 RGASMYRGVTRHHQHGRWQARI GRVAGNKDLYLGTFTQEEAAEAYDIAAIKFRGLNAV
 NFDISRYDVKSIASCNLPVGGMLPKPSPATAAADKTVDLSPSDSPSLTTPSLTFNVATPV
 NDHGGTFYHTGIPIKPDADHYWSNIFGFQANPKAEMRPLANFGSDLHNPSPGYAIMPMV
 QEGENNFGGSFVSGDGYNNHSAASNPVSAIPLSSTTTMSNGNEGYGGINWINNNISSY
 QTAKSNLSVLHTPVFGL*

>G170 (1..1107)

ATGGGGATGAAGAAGGTGAAGCTATCTTTGATAGCTAATGAAAGATCAAGGAAAACATCC
 TTCATAAAGAGGAAAGACGGGATTTTTAAGAACTCCACGAGTTGTCAACTCTGTGTGGT
 GTCCAAGCTTGTGCTCTCATCTACAGTCCATTCATACCGGTTCCAGAGTCATGGCCGTCA
 AGGGAAGGTGCTAAAAAGGTGGCTTCAAGGTTTCTGGAGATGCCGCCGACAGCCCGAACC
 AAGAAGATGATGGATCAAGAGACTTACCTTATGGAGAGGATTACCAAAGCAAAAGAGCAA
 CTAAGAACCTGGCTGCTGAGAACCGAGAGTTACAGGTTAGACGATTTATGTTTGATTGT
 GTTGAAGGCAAAATGTCCAGTATCATTATGATGCAAAAGACCTTCAAGATTTGCAATCT
 TGTATAAATCTATATCTCGATCAGCTTAACGGAAGGATCGAGTCCATTAAAGAAAATGGT
 GAGTCGTTGTTGTCTTCCGTCTCTCCTTTTCTTACTAGAAATTGGTGTGACGAAATTGGT
 GATGAGTCATTTTCCGACTCTCCTATTATGCTACAACCTGGGGTTGTAGATACTCTTAAT
 GCTACCAATCCTCATGTTCTTACGGGCGATATGACTCCTTTTCTTGATGCGGACGCAACT
 GCGGTAACGTGCTTCCAGTAGATTTTTTGATCATATTCCATATGAAAATATGAATATGAGT
 CAAAATCTGCATGAACCGTTTCAACACCTTGTTCCTACTAACGTTTGTGATTTTTTTCAA
 AATCAGAATATGAATCAGGTTCAATACCAGGCTCCTAATAATCTGTTTAATCAGATTCAA
 CGAGAATTCTACAACATAAATTTGAATCTGAATTTGAATCTGAATTCGAATCAGTATCTG
 AATCAACAACAATCATTCATGAATCCGATGGTGAACAACATATGAATCATGTTGGAGGG
 CGTGAAAGCATTCTTTTCGTGGACGGAACTGCTACAACCTACCATCAACTACCATCCAAT
 CAACTACCAGCCGTTGATCATGCTTCCACCAGTTACATGCCTTCCACCACCGGTGTCTAT
 GATCCTTACATCAACAATAATCTCTAA

>G170 Amino Acid Sequence (domain in aa coordinates: 2-57)

MGMKKVKLSLIANERSRKTSFIKRDGIFKKLHELSTLCGVQACALIYSPFIPVSWPS
 REGAKKVASRFLEMPPTARTKKMDQETYLMERITKAKEQLKNLAAENRELQVRRFMFDC
 VEGKMSQYHYDAKDLQDLQSCINLYLDQLNGRIESIKENGESLLSSVSPFPTRIGVDEIG
 DESFSDSPIHATTGVVDLTNATNPHVLTGDMTPFLDADATAVTASSRFFDHIPYENMMS
 QNLHEPFOHLVPTNVCDFFQONQNMNQVQYQAPNNLFNQIQREFYNINLNLNLNSNQYL
 NQQQSFMNPMVEQHMHVGGRESIPFVDGNCYNHQLPSNQLPAVDHASTSYMPSTTGVY
 DPYINNNL*

>G1768 (185..1426)

CTTCCTTTTGTCTCAGCTGCGAGCTTTGGTTGGATCTCTCACTTGCAAAACCAAATCCCT
 TATCGACTTCCACCGAAAGATCACTTCTTAACCTACACAAGGTGTTTGTATGAAGATCA
 GATAAATAAAAGGTCATTTGAGGATAATGGTTGATGTTCAAAGATTCTTACTTGCTTATT
 TGTGATGGACAATGTAAGAGGTTCAATAATGTTGCAGCCACTGCCAGAGATAGCTGAGAG
 TATCGATGATGCTATCTGCCATGAACCTCTCCATGTGGCCTGATGATGCTAAAGATTTGTT
 ATTGATAGTGGAGGCAATATCAAGGGGAGACTTGAAGTTGGTACTTGTTGCTTGTGAAA
 AGCTGTTTCTGAGAATAATCTTCTAATGGCAGGATGGTGTATGGGTGAGTTGCCGCGTAT
 GGTTCGATTTCTGGTGAGCCAATCCAGAGATTGGGAGCTTATATGTTAGAAGGGCTTGT
 TGCTAGGCTTGCTGCTTCTGTTAGTTCGATATATAAGTCTCTCCAGTCCAGAGAACCAGA
 GAGTTATGAATTTTATCTTATGTGTATGTTCTGCATGAGGTTTGTCCATATTTCAAGTT

TGGATACATGTCAGCGAATGGTGCGATTGCAGAAGCAATGAAGGATGAAGAGAGGATTCA
 CATTATTGACTTCCAAATTGGACAAGGGAGCCAGTGGATAGCACTTATCCAGGCTTTTGC
 AGCTAGGCCTGGTGGGGCTCCAAATATTCGAATTACCGGAGTTGGTGATGGATCTGTCTT
 GGTACAGTCAAGAAGAGACTAGAGAACTTGCAAAGAAGTTTGATGTTCCATTACAGGTT
 CAATGCGGTTTCAAGGCCAAGTTGTGAAGTTGAAGTGGAAAATCTTGATGTCCGAGATGG
 CGAAGCCCTTGGAGTGAACCTTTGCTTACATGCTGCATCATTGGCCAGATGAGAGTGTAAG
 CATGAAAAACACAGGGACCGGTTGCTGAGGATGGTGAAGAGTCTATCACCTAAAGTAGT
 CACTCTTGGAACAAGAATGCAACACGAACACTTCCCCTTCTCTCCTAGGTTCTTGA
 GACATTAAGTTATTACACGGCAATGTTTCAATCTATCGATGTTATGCTTCCGAGAAATCA
 CAAGGAAAGGATCAATATCGAGCAGCACTGCATGGCAAGGGATGTCGTCAACATCATAGC
 TTGTGAAGGAGCCGAGAGGATCGAAAGACACGAGCTTCTCGGGAAATGGAAGTCAAGGTT
 TTCCATGGCGGGTTTGTAGCCATACCCCTTGAGCTCAATCATTTCAGCCACCATTAGAGC
 CCTCTTGAGAGATTACAGCAACGGGTATGCGATTGAAGAAAGAGATGGTGCTCTGTACCT
 TGGTTGGATGGACCGAATCTTGGTCTCATCTTGTGCATGGAAGTGAAGAATAAACGTCT
 CCAAGAATGTAATGCAAAAGACAGAACCTGGAAGTAATAGATAGTTTGTCTCATAACCAT
 TAATAAGGTTGAATCAAATCATATACATCCCCATGCTACAACCTATTACACAGGCTCCATC
 AACAAAGAAGGGCTCTTGTGTGTACCTTCTCTTCTGTAACCTTTATTTGAACCAAT
 GGAAGTGGTTACAT

>G1768 Amino Acid Sequence (domain in AA coordinates: 54-413)

MDNVRGSI MLQPLPEIAESIDDAICHELSMWPDDAKDLLLIVEAISRGDLKLVLVACAKA
 VSENNLLMARWCMGELRGMVSI SGEPIQRLGAYMLEGLVARLAASGSSYKSLQSREPES
 YEFLSYVYVLHEVCYPYFKFGYMSANGAIAEAMKDEERIHIIDFQIGQGSQWIALIQAFAA
 RPGGAPNIRITGVGDGSLVLTVKKRLEKLAKKFDVPRFNAVSRPSCVEVENLDVRDGE
 ALGVNFAYMLHHLDPDESVMENHRDRLRMVKLSLSPKVVTLEQECNTNTSPFLPRFLET
 LSYYTAMFESIDVMLPRNHKERINIEQHCMARDVNNIACEGAERIERHELLGKWKSRFS
 MAGFEPYPLSSIISATIRALLRDYSNGYAIEERD GALYLGWMDRILVSSCAWK*

>G185 (77..988)

ATGCAAAAATAAACATAGTAACAATACTTTAAACTATTTACACCACCTTTAATCTTATTCT
 CCACTCTTTGAACGTAATGGAGAAGAACCATAGTAGTGGAGAGTGGGAGAAGATGAAGAA
 CGAGATCAACGAGCTAATGATAGAAGGAAGAGACTATGCACACCAGTTTGGATCAGCTTC
 ATCTCAAGAAACACGTGAACATTTAGCCAAAAAGATTCTTCAATCTTACCACAAGTCTCT
 CACCATCATGAACACTCCGGCGAAGTTGACCAAGTTTCTCAGGGTGGAGGAAGCCCCAA
 GAGCGATGATTCCGATCAAGAACCACCTTGTTCATCAAGAGTTTCAAGAAGTCAATGCCAAG
 GTGGAGTTCAAAAGTCAGAATTGCCCTGGAGCTGGTGTGATAGAACGCTGGACGATGG
 ATTCAAGTTGGAGAAAGTACGGCCAGAAGGATATTCTCGGAGCCAAATTTCCAAGAGGATA
 CTATAGATGCACGTATAGAAAGTCTCAAGGATGTGAAGCCACTAAACAAGTCCAAAGATC
 TGATGAAAATCAGATGCTCTTGAGATCAGTTACCGAGGAATACATTCTTGCTCTCAAGC
 TGCAATGTCCGTACAACATGCCGATACAAAACCTCGAACCAGAACCCAGACCCAAGAACA
 CGGAAATCTTGACATGGTAAAGGAAAGTGTAGACAACTACAATCACCAGCACATTTGCA
 TCACAACCTTCACTATCCATTGTCTATCTACCCCAATCTAGAGAATAACAATGCCTATAT
 GCTTCAAATGCGAGATCAAAACATCGAATATTTTGGATCTACGAGCTTCTCTAGTGATCT
 AGGAACTAGTATCAACTACAATTTTCCAGCATCTGGCTCGGCTTCTCACTCAGCATCAAA
 CTCTCCGTCCACCGTCCCTTTGGAATCCCCGTTTGAAGCTATGATCCAAATCATCCATA
 TGGAGGATTTGGTGGGTTCTATTCTTAGTTATCTACTTAAGGGAGGGACGGAACCTTTT
 CATGACCTCTTGATTAAAGAGAGAGTTTTTATAATAGCTAATCAATTTCTATTCAAATA
 TCCGAGTTTTTTTTCTAATCATGTTTATCAATTGTCTTATTACAGAAGGCTTATTTTTCAG
 GTCTATGTTGAAATAAATGGATTGTACTCGTAGGTATGATCCTTGTATTCTAAAAAAA
 AAAAA

>G185 Amino Acid Sequence (domain in AA coordinates: 113-172)

MEKNHSSGEWEKMKNEINELMIEGRDYAHQFGSASSQETREHLAKKILQSYHKSLTIMNY
 SGELDQVSQGGGSPKSDSDQEPLVIKSSKSMRWSKSVRIAPGAGVDRITLDDGFSWRK
 YGQKDILGAKFPGRYYRCTYRKSQGCATKQVQRSDENQMLLEISYRGIHSCSQANVGT
 TMPIONLEPNQTQEHGNLDMVKESVDNYNHQAHLHNLHYPLSSTPNLENNAYMLQMRD
 QNIEYFGSTSFSDDLGT SINYNFPASGSASHSASNSPSTVPLESPFESYDPNHPYGGFGG
 FYS*

>G1931 (5..592)

ATCAATGGAAGGGGTTGACAACACAAATCCTATGTTAACCTAGAGAAGGCGAAAACAA

CAATCCTTTTCTTCTTAGATGACAAAACATTAATGATGATGGCTCCTTCGTTAATCTT
 TTCGGGCGATGTAGGTCCATCTTCTTCTTGTACTCCAGCAGGTTATCATCTATCTGC
 TCAGCTGGAGAACTTTTCGAGGAGGTGGAGGAGAGATGGGAGGATTAGTGAGTAATAATAG
 CAATAATAGTGATCATAATAAGAATTGCAACAAAGGAAAAGGGAAGAGAAGCTTTGGCAAT
 GCAGAGGATAGCTTTTCATACAAGGAGTGATGATGATGTTCTTGATGATGGTTATCGTTG
 GCGAAAGTACGGTCAGAAATCTGTCAAGAACAATGCTCATCCCAGGAGCTATTATAGATG
 TACATACCACACATGCAACGTGAAGAAACAAGTGCAAGACTGGCAAAAGATCCAAACGT
 TGTGCTAACAACTACGAAGGTGTTTATCATCCCTTGAGAGAGCTCATGGAGACTCT
 TAGCCCTCTCCTTAGGCAACTTCAGTTCCTCTCAAGAGTTTCTGATCTGTAATTATTGAA
 TGTTAATTAGTGGTGAATACATTAATTATGCTTTAATCTCTCCATTGACCCTCAATC
 >G1931 Amino Acid Sequence (domain in AA coordinates: 114-170)
 MEGVDNTNPLMTLEEGENNNPFSSLDKTLMMAPSLIFSGDVGPPSSSSCTPAGYHLSAQ
 LENFRGGGGEMGGLVSNNSNNSDNHKNKNGKGRITLAMQRIAFHTRSDDDVLDGGRWR
 KYGQKSVKNAHPRSYRCTYHTCNVKKQVQRLAKDPNVVVTTYEGVHNHPCEKLMETLS
 PLLRQLQFLSRVSDL*

>G2543 (1..2169)

ATGAGTTTCGTCGTCGGCGTCGGCGGAAGTGGTAGTGGAAGCGGCGGAGACGGTGGTGGT
 AGTCATCATCACGACGGCTCTGAACTGATAGGAAGAAGAAACGTTACCATCGTCACACC
 GCTCAACAGATTCAACGCCTTGAATCGAGTTTCAAGGAGTGTCTCATCCAGATGAGAAA
 CAGAGGAACCAGCTTAGCAGAGAAATGGGTTTGGCTCCAAGACAAATCAAGTTCTGGTTT
 CAGAACAGAAGAAGTCAAGCTTAAGGCTCAACATGAGAGAGCAGATAATAGTGCATAAAG
 GCAGAGAATGATAAAATTGCTTGGGAAAACATTGCTATTAGAGAAGCTCTCAAGCATGCT
 ATATGTCCTAACTGTGGAGGTCCTCTGTTAGTGAAGATCCTTACTTTGATGAACAAAAG
 CTTTCGGATTGAAAATGCACACCTTAGAGAAGAGCTTGAAAGAATGTCTACCATGTCATCA
 AAGTACATGGGAAGACCGATATCGCAACTCTCTACGCTACATCCAATGCACATCTCACC
 TTGGATTGTCAATGACTAGTTTAACTGGTTGTGGACCTTTTGGTCATGGTCTTCACTC
 GATTTTGATCTTCTTCCAGGAAGTTCTATGGCTGTTGGTCTAATAATAATCTGCAATCT
 CAGCCTAACTTGGCTATATCAGACATGGATAAGCCTATTATGACCGGCATTGCTTTGACT
 GCAATGGAAGAATTGCTCAGGCTTCTTCAAGACAAATGAACCTCTATGGACAAGAACAGAT
 GGCTGCAGAGACATTCTCAATCTTGGTAGCTATGAGAATGTTTTCCCAAGATCAAGTAAC
 CGAGGGAAGAACCAGAACTTTCGAGTCGAAGCATCAAGGTCTTCTGGTATTGTCTTCATG
 AATGCTATGGCACTTGTGACATGTTTATGGATTGTGTCAAGTGGACAGAAGCTTTTCCC
 TCTATCATTCAGCTTCTTAAAACACTTGCAAGTATTTCTTCAAGAAATGGGAGGTACCCAT
 GAGGGTGCATTGCATTGTTGTATGAAGAAATGGAAGTGCTTTTCGCCTTTAGTAGCAACA
 CGCGAATTCTGCGAGCTACGCTATTGTCAACAGACTGAACAAGGAAGCTGGATAGTTGTA
 AACGTCTCATATGATCTTCTCAGTTTGTCTTCTCACTCTCAGTCTTATAGATTTCCATCT
 GGATGCTTGATTACAGGATATGCCAATGGATATCCAAGGTTACTTGGGTGAACATATT
 GAAACTGAAGAAAAGAACTGGTTTATGAGCTATACAGAGAGATTATTCAGAGGGGATT
 GCTTTTGGGGCTGATCGTTGGGTTACCACTCTCCAGAGAATGTGTGAAGATTGCTTCT
 CTATCGGTACCAAGCTTCTCATCTCGTGATCTCGGTGGAGTGATTCTATCACCGGAAGGG
 AAGAGAAGCATGATGAGACTTGCTCAGAGGATGATCAGCAACTACTGTTAAGTGTGAGC
 AGATCCAACAACACACGCTCAACCGTTGTTTTCGGAAGTGAACGAAGTTGGAATCCGTGTG
 ACTGCACATAAGAGCCCTGAACCAACCGGCACAGTCTTATGTGCAGCCACCCTTTCTGG
 CTTCCTCAATTCTCCTCAAAATGTCTTCAATTTCTCAAGACGAAAGAACCCGTCCTCAG
 TGGGATGTTCTTTCAAACGGAACGCAGTGCAAGAAGTTGCTCACATCTCAAACGGATCA
 CATCTTGGAACTGCATATCGGTTCTACGTGGATCCAATGCAACACATAGCAACAACATG
 CTTATTCTGCAAGAAAGCTCAACAGACTCATCAGGAGCATTTGTGGTCTACAGTCCAGTG
 GATTTAGCAGCATTTGAACATCGCAATGAGCGGTGAAGATCCTTCTTATATTCTCTCTTG
 TCCTCAGGTTTTCACAACTCACCAGATGGAATGGCTCAAACCTGAAACAAGGAGGAGCC
 TCGACGAGCTCAGGACGGGCATCAGCTAGCGGTTCTGTTGATAACGGTTGGGTTTCAGATA
 ATGGTAAGCAATTTACCGACGGCAAACTGAATATGGAGTGGTGGAAACGGTTAATAAC
 CTGATAGGAACAACGTGATCAAAATTAACCGCCTTGAGCGGTCCTACAGCTTCAACT
 ACAGCTTGA

>G2543 Amino Acid Sequence (domain in AA coordinates: 31-91)
 MSFVVGVGSGSGSGGGGSHHHDGSETDRKKRYHRHTAQQIQRLSSFKPCPHPEK
 QRNQLSRELGLAPRQIKFWFQNRRTQLKAQHERADNSALKAEENDKIRCENIAIREALKHA
 ICPNCGGPPVSEDPYFDEQKLRIENAHLEELERMSTIASKYMGRPISQLSTLHPMHISP

LDLSMTSLTGCGPFGHGPSLDFDLLPGSSMAVGPNNNLQSQPNLAISMDKPIMTGIALT
 AMEELLRLQLTNEPLWTRTDGCRDILNLGSENVFPRSSNRGKNQNFVEASRSSGIVFM
 NAMALVDMFMDCVKWTLPFSIIAASKTLAVISSGMGGTHEGALHLLYEEMEVLSPLVAT
 REFCELRQCQTEQGSWIVNVSYDLPQFVSHSQSYRFPSCGLIQDMPNGYSKVTWVEHI
 ETEEKELVHELYREIIHRGIAFGADRWVTTLQRMCFASLSVPASSSRDLGGVILSPEG
 KRSMMLAQRMISNYCLSVSRNNRSTVSELNEVGIRVTAHKSPEPNGTVLCAATTFW
 LPNSPQNVFNFLKDERTRPQWDVLSNGNAVQEVAHISNGSHPGNCISVLRGSNATHSNM
 LILQESSTDSSGAFVVYSPVDLAALNIAMSGEDPSYIPLSSGFTISP DGNGSNSEQGGA
 STSSGRASASGLITVGFQIMVSNLPTAKLNMESVETVNNLIGTTVHQIKTALSGPTAST
 TA*

>G264 (30..1430)

CTTGTACCAGTTTCTGATTAGATTCAACAATGAACGGCGCATTAGGTAACCTCCTCCGCCT
 CCGTTAGCGGCGGAGAAGGAGCCGGAGGACGCGCCTTCTTGGTGAAAACCTACGAGA
 TGGTCGACGATTCAACAGGACGATCGTATCGTGGAGCGCTAACACAACAGCTTCA
 TCGTTTGGAAATCATGCCGAATTTTCACGCCTCCTTCTTCCAACCTACTTCAAACACAATA
 ACTTCTCTTCTTCAATTCGTGCTCAATACCTATGGGTTTAGGAAGATTGATCCAGAGA
 GGTGGGAGTTTTTGAATGATGATTTTATTAAGGATCAGAAGCATCTTCTCAAGAATATAC
 ATAGAAGGAAACCTATACACAGCCACAGTCATCCACCTGCTTCGTCGACTGATCAAGAAA
 GAGCAGTGTTCAGAGCAAATGGACAAGCTTTCACGTGAGAAAGCTGCAATTGAAGCTA
 AGCTTTTAAAGTTCAAACAACAGAAGGTTGTAGCAAAGCATCAGTTTGAAGAAATGACTG
 AGCATGTTGATGATATGGAGAATAGGCAGAAGAAGCTGCTGAATTTTTTGGAACTGCCA
 TTCGGAATCCTACTTTTGTGAAGAAATTTGGTAAGAAAGTCGAGCAGTTGGATAATTCAG
 CTTACAACAAAAAGCGAAGGCTCCCTGAAGTTGAGCAATCAAAGCCACCTTCAGAAGATT
 CTCATCTGGATAATAGTAGTGGTAGCTCGAGACGCGAGTCTGGAAACATTTTTTCATCAAA
 ATTTCTCTAATAAATTGCGACTAGAGCTTTCTCCAGCTGATTGAGATATGAACATGGTTT
 CACACAGTATACAAAGTTCCAATGAAGAAGGTGCGAGTCCCAAAGGGATACTGTCAGGAG
 GTGATCCAAATACTACACTAACAAAAAGAGAAGGCCTACCATTTGCACCTGAAGCTCTAG
 AGCTTTCGCGGATACCGGGACATGCCCGAGGAGATTACTGTTAAATGATAATACAAGGGTGG
 AGACCTTTCAGCAGAGGCTAACTTCTTCAGAGGAGACTGATGGTAGCTTTTTCATGTCATT
 TAAATCTAACCTTGGCTTCTGCTCCGTTACCGGACAAAACAGCTTCACAGATAGCTAAGA
 CGACTCTTAAAAGTCAGGAGTTAACTTTAACTCAATAGAAACAAGTGCAAGTGAGAAAA
 ATCGGGGTAGACAAGAGATTGCGAGTTGGAGGTAGCCAAGCAAATGCGAGCTCCTCCAGCAA
 GAGTGAATGATGTATTCTGGGAACAGTTCCCTAACAGAAAGGCCAGGGTCTTCAGATAATG
 AGGAGGCAAGTTGACTTATAGAGGTAACCCATACGAAGAGCAAGAGGAGAAAAGAAACG
 GGAGTATGATGTTACGTAATACAAAGAATATCGAGCAGCTGACCTTATAAACTATTTGGA
 CGGTTACATCAACGAGAGTACGAAGTGGGTTTGGTAAGAAAGTATGGGTGAGTAAGTAA
 TGAACATTGGACTGAAAAAGCGTAAGTAGCTTTGTTGTAAACACTTGCGTCTCTGTCTA
 CACAAGTAATTTGACTGTAAATGTAAGTGTACAGGATTTAAATTGAATAAGCA

>G264 Amino Acid Sequence (domain in AA coordinates: 24-114)

MNGALGNSSASVSGGEGAGGPAPFLVKTYEMVDDSSDQIVSWANNNSFIVWNHAEFSR
 LLLPTYFKHNNFSSFIRQLNTYGFRKIDPERWEFLNDDFIKDQKHLLKNIHRRKPIHSHS
 HPPASSTDQERAVLQEQMDKLSREKAAIEAKLLKFKQKQVVAKHQFEEMTEHVDDMENRQ
 KKLNLNLETAIRNPTFVKNFQKKEQLDISAYNKKRRLPEVEQSKPPSEDSHLDNSSGSS
 RRESGNIFHQNFNKLRLLELSPADSDNMVSHSIQSSNEEGASPKGILSGGDPNTTLTKR
 EGLPFAPEALELADTGTCPRRLLLNNDNTRVETLQQLRSTSEETDGSF SCHLNLTLASAPL
 PDKTASQIAKTTLLKSQELNFNSIETSASEKNRGRQEIIVGGSQANAAPPARVNDVFWEQF
 LTERPGSSDNEEASSTYRGNPYEEQEEKRNKSGMMLRNTKNIEQLTL*

>G32 (101..736)

AACACACATTCCTCTCTTCTTCACTAGAAAAAGATAGATATATCGGACATTTATTG
 ATCTGTGTATGCAAGGTATAGTATCATTTATTAGAAAGATGAACACAACATCATCAAA
 GAGCAAGAAGAAGCAAGACGATCAGGTTGGTACAGGTTTCTTGGGGTGAGAAGAAGGCC
 TTGGGGAAGATACGACGCTGAGATTAGAGACCCAACTACGAAGGAGCGTCACTGGCTTGG
 CACTTTCGATACGGCGGAAGAAGCTGCCTTGGCCTACGATAGAGCTGCTCGGTCCATGCG
 TGGCACACGTGCCAGAACCAACTTTGTTTACTCAGACATGCCTCCTTCCTCATCCGTCAC
 CTCCATTGTTTCTCTGACGATCCTCCTCCTCCTCCACCTCCTCCTGCTCCTCCTAGCAA
 TGATCCTGTGCTGATTACATGATGATGTTTAAACCAATACTCATCCACTGACTCGCCAATGCT
 TCAGCCTCATTGTGATCAAGTGGACAGTTACATGTTTGGTGGCTCTCAATCTTCGAATTC

TTATTGCTATTCTAATGACAGTAGTAATGAGCTGCCTCCTCTCCCGAGCGACTTGTCGAA
TTCGTGTTATAGCCAACCACAGTGGACCTGGACCGGTGACGACTACTCGTCTGAGTACGT
ACATAGTCCAATGTTTCAGCAGAATGCCTCCGGTTTCTGACTCTTTCCCTCAAGGTTTCAA
CTACTTTGGCTCCTAATCTTTCTCATCGTCCATATTTAATACCTTCCTCATTTGTACCT
TTTCCTTCTTCTCTTTTGGGTTTATCTATGTTTCGCGCTCCTTGATCTCTGCCTATG
TGATCAAAGTGACTGTTTGTTCATTAGTTTTTCAATAACAAGTTATCATTTGTATCTTGAA
AAAAAAAAAAAA

>G32 Amino Acid Sequence (domain in aa coordinates: 17-84)
MNTTSSKSKKKQDDQVGTRFLGVRRRPWGRYAAEIRDPTTKERHWLGTFTDABEALAYD
RAARSMRGTRARTNFVYSMDPPSSSVTSIVSPDDPPPPPPPPAPPSPNDPVDYMMMFNQYS
STDSPMLQPHCDQVDSYMFQGSQSSNSYCYNSDSSNELPPLPSDLSNSCYSQPQWTWTGD
DYSSEYVHSPMFSRMPVSDSPFQGFNYFGS*

>G436 (1..2157)

ATGGATTTTACTCGCGATGACAACCTCAAGTGATGAACGGGAAAATGATGTAGACGCCAAC
ACCAACAACCGTCACGAGAAGAAGGGTTACCATCGCCACACTAATGAACAAATTCATAGG
CTTGAAACGTATTTCAAGGAATGTCCTCATCCAGACGAATTTAGCGAGCTCTGTTGGGT
GAAGAACTGAATCTGAAACCAAAACAAATCAAATTTTGGTTTCAAAACAAAAGAACTCAA
GCTAAGAGTCACAATGAAAAAGCAGACAATGCAGCGCTTAGGGCAGAAAATATTAAGATT
AGACGTGAGAACGAATCAATGGAAGATGCACCTGAATAATGTGGTTTGGCCTCCATGTGGT
GGTCGTGGTCTCGGAGAGAAGACCAACTTCGACATCTCCAAAACCTCCGTGCACAAAAC
GCTTATCTCAAAGATGAGTATGAAAGAGTCTCAAACCTACCTAAAACAGTACGGAGGTCAC
TCAATGCTAATACGTCGAGGCCACACCTATCTCCATGGTCCATCAAACCATGCATCAACG
TCCAAGAACCGTCCAGCATTGTACGGAACCTCTTCTAACCGTCTCCCGAGCCTTCAAGC
ATATTTAGAGGACCATACTCCTGGAACATGAACACCACCGACCGCTCAGCCGCGA
AAGCCGCTGGAATGCAGAAATTTCAACCACTATCTCAACTGGAGAAAATGCAATGTTG
GAAGCAGCGGAAAAGCGGTGTGAGAGGTTTTGAGCCTCATTCAAATGGATGATACAATG
TGGAAAAAGTCGTCTATTGATGATAGGCTCGTCATTGATCCAGGGCTCTATGAGAAATAT
TTTACTAAGACTAACACAAATGGTCGTCTGAGTCTTCTAAAGATGTCGTGGTGGTTCAA
ATGGATGCTGGAACCTTGATCGACATCTTCTTAACTGCGGAGAAATGGGCGAGGCTTTTT
CCAACAAATGTGTAACGAAGCTAAAACGATTACGCTCTGGATTCCGTTGACCATCGAGGA
AAAACCTTTCTCAAGAGTGATTATGAGCAACTGCACATACTGTCAACATTGGTGCCACCG
AGGGAATTTATGATCCTAAGGACTTGCCAACAAATGGAAGACAATGTCTGGATGATTGCT
GATGTGTGCTGTCATCTCCCAACATTGAGTTTGATCTTTCTGTTTCCCATTTGCACCAA
CGTCCCTCAGGTGTGCTCATTCAAGCCTTGCCCCACGGCTTCTCTAAGGTGACGTGGATA
GAGCATGTGGTAGTGAATGATAATAGAGTGCAGGCCACATAAGCTTTACAGAGACCTCTTA
TACGGCGGCTTTGGCTACGGAGCTCGACGTTGGACCGTTACTCTTGAGAGGACGTGTGAG
AGGCTGATTTTCTCCACCTCCGTCCCTGCCCTTGCCCAACAATGACAATCCCGGAGTTGTG
CAAACAATACGAGGCAGAAATAGCGTAATGCATTTGGGAGAAAGATGTTGAGGAACCTT
GCATGGATGATGAAAATGGTTTAAACAACTCGACTTCTCGCCACAGTCTGAAACTAACAAC
AGCGGAATTAGGATTGGGGTGCGGATAAACAATGAGGCGGGTCAACCGCCCGGTCTCATT
GTCTGTGCTGGTTTCATCTTTATCCCTCCCTCTCCCTCCTGTCCAAGTGACGATTTCTT
AAGAATCTGGAGGTTCTGACCAAGTGGGACGTTCTGTGCCATGGGAATCCAGCGACTGAG
GCTGCTCGTTTCTGTCACCGGATCAAACCCAAGGAACACTGTGTCTTTTCTCGAGCCTTCA
ATTAGGGATATTAATACTAAGCTAATGATACTCCAAGATAGCTTCAAAGATGCATTGGGA
GGAATGGTGGCCTACGCTCCAATGGATCTAAACACCGCCTGCGCTGCCATTTTCAGGCGAT
ATCGATCCTACCACCATTTCAATCCTCCCTTCCGGTTTTATGATCTCCCGTGACGGCCGT
CCTTCCGAGGCGAAGCCGAGGGTGCGAGCTATACACTCCTCACCCTGGCTTTCCAGATC
CTTGCTCTCCGGTCCSAGTTACTCTCTGATACCAACCTGGAAGTTCTGTCACCCACAGTC
AATACCTTGATTAGCTCCACCGTTCAAAGGATCAAAGCCATGCTCAAGTCCGAATGA

>G436 Amino Acid Sequence (domain in AA coordinates: 22-85)
MDFTRDDNSSDERENDVDANTNNRHEKKGYHRHTNEQIHRLETYFKECPHPDEFQRRLLG
EELNLKPKQIKFWFQNKRTQAKSHNEKADNAALRAENIKIRRENESMEDALNNVCPPCG
GRGPGREDQLRHLQKLRAQNAYLKDEYERVSNYLKQYGGHSMHNVEATPYLHGSPNHAST
SKNRPALYGTSSNRLPEPSSIFRGPYTRGNMNTAPPQPRKPLEMQNFQPLSQLEKIAML
EAAEKAVSEVLSLIQMDTMMWKSSIDRLVIDPGLYEKYFTKTNTNGRPESKDVVVVQ
MDAGNLIDIFLTAEKWARLFPTIVNEAKTIHVLDSVDHRGKTFSRVIYBQLHLISPLVPP
REFMILRTCQOIEDNVNMIADVSHLPNIEFDLSFPICTKRPSGVLIQALPHGFSKVTWI

EHVVVNDNRVRPHKLYRDLLYGGFGYGARRWTVTLERTCERLIFSTSVPALEPNNDNPGVV
QTIRGRNSVMHLGERMLRNFAMMMKMNKLDSPQSETNNSGIRIGVRINNEAGQPPGLI
VCAGSSLSLPLPPVQVYDFLKNLEVRHQWDVLCHGNPATEAARFVTGSNPRNTVSFLEPS
IRDINTKLMILQDSFKDALGGMVAYAPMDLNTACAAISGDIDPTTIPILPSGFMISRDGR
PSEGEAEGGSYTLTLTVAFQILVSGPSYSPDTNLEVSATTVNTLISSTVQRIKAMLKCE*
>G556 (50..1144)

CTTTTTTGAAGCCCTTTTGACACAAAAGACCAGAACAAAGTTGAAGAAATATGAATACAAC
CTCGACACATTTTGTTCACCGAGAAGGTTTGAAGTTTACGAGCCTCTCAACCAAATCGG
TATGTGGGAAGAAAGTTTCAAGAACAAATGGAGACATGTATACGCCTGGCTCTATCATAAT
CCCGACTAACGAAAAACCAGACAGCTTGTCTAGAGGATACTTCTCATGGGACAGAAGGAAC
TCCTCACAAAGTTTGACCAAGAGGCTTCCACATCTAGACATCCTGATAAGATACAGAGAAG
GCTAGCACAGAATCGAGAGGCAGCTAGGAAAAGTCGTTTGCAGCAAGAAAGCTTATGTTCA
GCAGCTAGAGACTAGCCGGTTAAAGCTAATTCATTTAGAGCAAGAACTCGATCGTGCTAG
ACAACAGGGTTTCTATGTGGGGAACGGAGTAGATACCAATGCTCTTAGTTTCTCAGATAA
CATGAGCTCAGGGATTGTTGCATTTGAGATGGAATATGGACATTGGGTGGAAGAACAGAA
CAGGCAAAATATGTGAACATAAGAACGGTTTTACATGGACAAGTTAGTGATATAGAGCTTCG
TTCTCTAGTCGAGAATGCCATGAAACATTACTTTCAACTCTTCCGAATGAAGTCAGCCGC
TGCAAAAAATCGATGTTTTCTATGTCATGTCCGGAATGTGGAAGAACTTCAGCAGAGCGGTT
TTTCTTGTGGATAGGCGGATTTAGACCCTCAGAGCTTCTCAAGGTTCTGTTACCGCATT
TGATCCTTTGACGGATCAACAACCTTTGGATGTATGTAATCTGAGGCAATCATGTCAACA
ATCAGAAGATGCGTTATCCCAAGGTATGGAGAACTGCAACATACATTAGCAGAGAGTGT
AGCAGCCGGGAAACTTGGTGAAGGAAGTTATATTCCTCAAATGACTTGTGCTATGGAGAG
ATTGGAGGCTTTGGTCAGCTTTGTAAATCAAGCTGATCATCTGAGACATGAGACATTGCA
ACAGATGCATCGGATCTTAACCACGCGACAAGCGGCTAGAGGTTTGTAGCATTAGGGGA
GTATTTCCAAAGGCTTCGAGCTTTGAGTTCGAGTTGGGCGGCTAGGCAACGTGAACCAAC
GTAATTAAGGTGTTTAGATGTCAAGAAAGGTTTGAGACCTTAACAATCAAGAAATGGAGTT
TGCTGGTGAAGTGGATTTTTGGGTCAAGAACAAAGAGCAATAACACAAGCTGCTGTGTGATG
ATGAATCTTGTCTTGCGGCTAAAGGAAATGTTTGAGGAAAGTTGTACATATGATCAGCAA
CGTAAAGTTTATAGCTTTTGTAGAACCAACTTTTCGATGGTTGTTCTTTTTTTTTTGTAT
GTAATATTATAGATAAGCTTGTGGTATATATGATTTTAATGTGACATTACGAACCTTGATT
TATAACCATGGTAAAAT

>G556 Amino Acid Sequence (domain in AA coordinates: 83-143)
MNTTSTHFVPPRRFEVYEPLNQIGMWEESEFKNNGDMYTPGSIIIPTNEKPDLSLSEDTSHG
TEGTPHKFDQEAESTSRHPDKIQRRLAQNREARKSRLRKKAYVQOLETSRLKLIHLEQEL
DRARQQGFYVGNVDNLSFSDNMSSGIVAFEMEYGHWWVEEQNRQICELRTVLHGQVSD
IELRSLVENAMKHYFQLFRMKSAKIDVFYVMSGMWKTSARFFLWIGGFRPSELLKVL
LPHFDPLTDQQLLDVCNLRQSCQSEDALSQGMELQHTLAESVAAGKLGEGSYIPQMT
AMERLEALVSFVNQADHLRHETLQQMHRILTTTQAARGLLALGEYFQRLRALSSSSWAARQ
REPT*

>G1420 (39..1238)

AAAGTATCATCTCATAGATTCCATCTTTTCTCTATTACATGGAGAAGAAAAAGAGAGG
ATCATCATCATCAACAACAACAACAACAAGGAGATCAAGAACACAGAGACAAAGA
TCGAGCAAGAACAAAGAACAAAGAAAAACAAGAAATCTCTCAAGCATCATCATCAA
ACATGGCGAATCTAGTTACGTCATCAGATCATCATCCGTTGGAGCTAGCTGGAATCTCT
CAAGCATCTTCGATACTTCATCTTTACCTTTTCCCTTATCTTATTTTGAAGATCACTCTT
CTAATAATCCTAATCTTTTCCCTAGACTTGCTCCGACAAGATCATCAGTTTGTCTTCTCT
CTAATTCCTCTCTTTTTCATTTCGATGCCTTTCTCTCTCCCAATAACAACAACAACCT
CTTTTTTTACGGATTGCCCCTTACCTCAAGCTGAGTCATCAGAAGTCGTGAACACAACAC
CGACTTCTCCAAACTCAACCTCAGTCTCATCTTCTCCCAACGAAGCTGCAATGATAACA
ACAGTGGTAAAGAGTTACTGTTAAAGATCAAGAAAGAGGAGATCAACAACAAGAGCAAA
AGGGTACTAAGCCACAGTTGAAGGCAAGAAAGAAATCAAAAGAAAGCTAGAGAAGCTA
GGTTTGCCTTTCTGACGAAGAGCGATATTGATAATCTTGACGACGGTTATAGGTGGAGAA
AATACGGCCAAAAAGCTGTCAAAAACAGTCCTTATCCAGAAGCTATTACCGTTGCACCA
CAGTGGGTTGCGGAGTGAAGAAGAGAGTGGAGAGATCTCCGATGATCCTTCGATCGTCA
TGACAACCTACGAAGGTCAGCATACCCATCTTTTCCCCATGACGCCACGTGGACACATCG
GAATGCTCAGGTCACCAATCCTAGACCAGGTGCAACCACCGGTCATCATCATCATTCT
CCATCCCTCAGCCACGTTACTTGCTGACTCAACATCACCAGCCCTACAACATGTACAACA

ACAACTCTCTAAGTATGATCAATAGAAGATCATCCGATGGCACTTTCGTAAATCCAGGTC
CATCATCATCATCCCCGGCTTTGGTTATGATATGTCTCAAGCTTCTACTTCAACTTCTT
CTTCCATTAGAGATCATGGATTGCTTCAAGATATCTTCTTCGCAGATCAGATCCGATA
CTATTAACACTCAAACCAATGAAGAGAATAAGAAATGAAGAAGTTTTTTTCCCCGGGGCA
ATTGTTTTTTTCTTTAGGCCGGATCCGGTAGGTAGGTTTCATGAGC

>G1420 Amino Acid Sequence (domain in AA coordinates: 221-280)
MEKKKEEDHHHQQQQQQQKEIKNTETKIEQEQQEQKEISQASSSSNMANLVTSSDHP
LELAGNLSIFDTSLLPFPYSYFEDHSSNNPNSFLDLLRQDHQFASSNSSSSFSFADFPL
PNNNNNTSFFTDLPLPQAESSEVVNTTPTSPNSTSVSSSSNEAANDNNSGKEVTVKDQEE
GDQQQEOKGTPQLKAKKKNQKKAREARFAFLTKSDIDNLDGYSRWKYGQKAVKNSPYP
RSYYRCTTVGCGVKRVERSSDDPSIVMTTYEGQHTHPFPMTPRGHIGMLTSPILDHGAT
TASSSSFSIPQPRYLLTQHHQPYNNMNNNSLSMINRRSSDGTFFVNPGPSFPGFGYDMS
QASTSTSSSIRDHGLLDILPSQIRSDTINTQTNEENKK*

>G1412 (115..1008)
CCCACGCGTCCGCCACGCGTCCGAAACAAAAACATATAATTTGGGTTTTTAGAGTTCGA
AACTTGAAATCTTTTTTTTTTTGGTTGCTGAGGAATCGAAGTAGAAGAGTATAAATGGGT
GTTAGAGAGAAAGATCCGTTAGCCAGTTGAGTTGCCACCAGGTTTTAGATTTTATCCG
ACAGATGAAGAGCTTCTTGTTCAGTATCTATGTCGGAAGTTGCAGGCTATCATTTCTCT
CTCCAGGTCATCGGAGACATCGATCTCTACAAGTTCGATCCTTGGGATTGCGCAAGTAAG
GCTTTGTTGGAGAGAAGGAATGGTATTTCTTTAGCCCCAAGAGATCGGAAATATCCGAAC
GGGTCAAGACCCAATAGAGTAGCCGGGTCGGGTATTGGAAAGCAACGGGTACTGACAAA
ATTATCACGGCGGATGCTCGTGTGCGGATTAAAAAGCTCTGGTCTTTACGCCGGA
AAAGCTCCCAAAGGCATAAAACCAACTGGATTATGCACGAGTATCGCTTAATAGAACAT
TCTCGTAGCCATGGAAGCTCCAAGTTGGATGATTGGGTGTTGTGTGCAATTTACAAGAAA
ACATCTGGATCTCAGAGACAAGCTGTACTCCTGTTCAAGCTTGTCTGTAAGAGCATAGC
ACGAATGGGTCGTCTATCTCTTTCATCACAGCTTGACGACGTTCTTGATTCTGTTCCCG
GAGATAAAAGACCAGTCTTTTAATCTTCTCGGATGAATTCGCTCAGGACGATTCTTAAC
GGGAACTTTGATTGGGCTAGCTTGGCAGGCTTTAATCCAATTCAGAGCTAGCTCCGACC
AATGGATTACCGAGTTACGGTGGTTACGATGCGTTTCGAGCGGCGGAAGGTGAGGCGGAG
AGTGGGCATGTGAATCGGCAGCAGAACTCGAGCGGTTGACTCAGAGTTTCGGGTACAGC
TCGAGTGGGTTTGGTGTTCGGGTCAAACATTGAGTTTAGGCAATGAGAGAGATGTGAA
GTTACTGATGGGTGAAAAAAGTAAAAAATACTGGAGATAGTAGAGTGGCAATTGATG
TAAATAATAGGGATTTATATGGGGCTTTTACCGATTGCGGTGAGGCTTAGGATTCCCCAAA
GGAAAAAGGCTCGACTGGGGACTAGTTTGATCCAACCTTGACGGCCCCCAATGTGTAATG
TTTCTCAACGGAGAGAAAAATAATGGTTACCAATATTTTCCAAAAAATAAATAAATAA
>G1412 Amino Acid Sequence (domain in AA coordinates: 17-159)
MGVREKDLPLAQLSLPPGFRFYPTDEELLVQYLCKRVAGYHFSLVQVIGDIDLYKFPDPLP
SKALFGEKEWYFFSPDRDKYPNGSRPNRVAGSGYWKATGTDKIIADGRRVGIKKALVFI
AGKAPKGTKNWIMHEYRLIEHSRSHGSSKLDWVLCRIYKKTSGSQRQAVTPVQACREE
HSTNGSSSSSSQLDDVLDSPFIKQDSFNLPRMNSLRITLNGNFDWASLAGLNPIPELA
PTNGLPSYGGYDAFRAEAGEAESGHVNROQNSSGLTQSFYSSSGFGVSGQTFEFRQ*

>G738 (1..885)
ATGGACCATCATCAGTATCATCATCATGATCAATACCAACATCAGATGATGACTAGTACT
ACAATAATTCCTATAACACCATCGTCACAACACAACCACCACCAACAACAACAATG
GATTCAACAACAGCAACAACATATGATAATGGATGACGAGAAGAAGTTGATGACGACAATG
AGCACTAGGCCGCAAGAACCAAGAACTGTCCAAGATGCAACTCAAGCAACACCAAGTTT
TGTTATTACAACAACACTAGCTTAGCACAGCCTAGGTACTTGTGTAAGTCTTGTGCGGAGA
TATTGGACTGAAGGTGGCTCTCTCCGTAACGTCCCGTAGGCGGAGGTTCTAGAAAGAAC
AAGAAGCTTCCATTTCCCTAATTCCTCTACTTCTTCTCCACCAAGAACCTCCCGGATCTC
AACCCTCCTTTCTGTTCTCACATCATCAGCTTCATCATCAAACCCTAGCAAGACGCATCAA
ACAATAATGACCTCAGCCTATCCTTCTCCTCCCTATGCAAGACAAGCGAGCTCAAGGG
CATTACGGTCATTTTCAGTGAGCAAGTTGTGACAGGAGGGCAGAACTGTCTTTTCCAAGCT
CCTATGGGAATGATTTCAGTTTCGTAAGAGTATGATCATGAGACCCCCAAAAGAATCTT
GGGTTTTTCATTAGACAGGAACGAGGAAGAGATTGGTAATCATGATAACTTCGTTGTTAAT
GAGGAAGGAAGTAAGATGATGATCCTTATGGAGATCATGAAGACCGTCAACAACATCAC
CATGTGAGACACGATGATGGTAATAAGAAGAGAGAAGGTGGTTCAAGCAATGAGCTATGG
AGCGGAATCATCCTAGGTGGTGTAGTGGTGGACCAACATGGTGA

>G738 Amino Acid Sequence (domain in aa coordinates: 351-393)
 MDHHQYHHHDQYQHMMSTNNNSYNTIVTTQPPPTTTTMDSTATTMIMDDEKKLMTTM
 STRPQEPNRCPRCNSSNTKFCYYNNYSLAQPRYLCKSCRRYWTEGGLRNVPVGGGSRKN
 KKLPPFNSSTSSSTKNLPLNPPFVFTSSASSSNPSKTHQNNNDLSLSFSSPMQDKRAQG
 HYGHFSEQVVTGGQNCFLQAPMGMIQFRQEYDHEHPKKNLGFSLDRNEEEIGNHDFVFN
 BEGSKMMPYGDHEDRQHHHVRHDDGNKKREGGSSNELWSGIILGGDSGGPTW*

>G2426 (1..1038)

ATGGGCAGATCGCCATGTTGTGATAAGGCCGGGTGAAGAAAGGCCCTTGACTCCAGAA
 GAGGATCAGAACTTTTGGCTTATATTGAAGAACATGGCCATGGAAGCTGGCGTTCTTTG
 CCTGAGAAAGCCGGTCTCCAAAGGTGTGGAAAGAGTTGCAGACTCAGATGGACTAACTAC
 CTAAGACCTGACATCAAGAGAGGCAAAATTCAGTGTACAAGAAGAACAAACCATCATTCAA
 CTCCACGCTCTCCTCGGAAACAGGTGGTCAGCGATTGCAACTCATTTACCAAAGAGGACA
 GACAACGAGATCAAGAACTACTGGAACACACACTTGAAGAAACGTCTGATCAAAATGGGG
 ATAGATCCAGTGACTCACAAGCACAAAAACGAGACTCTTTCTGCTTCCACAGGACAATCA
 AAGAACGCGAGCCAGCTTAGTCATATGGCTCAATGGGAGAGTGCAAGACTCGACGCTGAA
 GCAAGGCTAGCTAGAGAAATCAAAGCTTCTCCATTTACAGCATTACCAAACAATAACAAC
 CTTAACAATCAGCAGCTCCTCAACAACATTGCTTCACTCAAAAAACATCAACAACTGG
 ACTAAACCAAACCAAGGAAACGGAGACCAACAGCTTGAATCTCCGACATCGACGGTGACA
 TTCTCTGAGAATCTTCTGATGCCTTTAGGAATCCCTACGGATAGCAGCAGAAATAGAAAC
 AATAACAACAATGAGTCTCGGCGATGATTGAATTGGCCGTATCTTCGTCAACCTCCTCC
 GATGTGAGTCTGGTCAAAGAACATGAACACGACTGGATTAGGCAGATCAACTGTGGTAGT
 GGAGGAATAGGAGAAGGATTACGAGTCTATTGATCGGTGATTTCGGTCGGCCGGGGTTTA
 CCCACCGGGAAAAACGAAGCGGCGGCGTGGGGAATGAGAGTGAGTATAACTACTAT
 GAGGATAACAAGAATTACTTGAATAGCATTCTCAACTTGGTTGATTCTTCACCGTCCGAT
 TCCGCGACGATGTTCTGA

>G2426 Amino Acid Sequence (conserved domain in AA coordinates:14-114)

MGRSPCCDKAGLKKGPWTPEEDQKLLAYIEEHGHSWRSLEKAGLQRCGKSCRLRWNTY
 LRPDIKRGKFTVQEEQTIQLHALLGNRWSAIATHLPKRTDNEIKNYWNTHLKKRLIKMG
 IDPVTHKHKNETLSSSTGQSKNAATLSHMAQWESARLDAEARLARESLLHLQHYQNNNN
 LNKSAAPQQHCFTQKTSTNWTNPKPNQNGDQQLSPTSTVTFSENLLMPLGIPTDSSNRN
 NNNNESSAMIELAVSSSTSDVSLVKEHEHDWIRQINCGSGGIGEGFTSLIGDSVGRGL
 PTGKNEATAGVGNESYNYEDNKNYWNISILNLVDSSPSDSATMF*

>G1524 (1..825)

ATGGGGAGAACTAAGGAGCAGGCAACATTAACCTCGGTATCCACCCTGTCTTAGGAATCCT
 GCTAAATTCAATGATATAAACAAGCACTCCAGGAAAAAGGATATGGTAAGGCTCTGAAA
 AGAAAACCTTGACCGGTGTGACATGCCCTGTCTGTCTTGAGGTTCTCACAACCTCGGT
 GTCCTCCTTTGTTTCTTACCACAAAGGATGCCGTCCGTACATGTGTGCCACGGGAAAC
 CGTTTCTCAAATTGTCTAGAGCAGTACAAAAAGGCATATGCCAAGGATGAGAAAAGTGAC
 AAACCGCCAGAGCTATTGTGCCCGCTTTGTAGGGGTGAGTGAAAGGCTGGACCGTTGTG
 GAAAAGGAACGTAAGTATCTGAATTCTAAGAAAAGGTCATGATGAACGACGAGTGTG
 TTTTATGGAAGCTATAGACAGCTCAAGAAGCATGTTAAGGAGAACCATCCGAGAGCCAAG
 CCAAGAGCCATAGACCTGTGCTGGAGGCGAAATGGAAGAAGCTTGAGGTTGAGAGGGAG
 AGGAGTGATGTAATCAGCACAGTCATGTCTCAACACCTGGGGCTATGGTATTTGGAGAC
 TATGTGATTGAGCCATACAATGGTTATGATCATCAAGATGACAGTGACGATTACAGTGAT
 TCGTCCGATGACGAAATGGAAGGTGGGGTATTCGAGCTTGAGCATTTCGACCTGGGCCGT
 CTTCAACCGCGTTTCGGCTGCCATCTCAAGCCGGGGAATTCGCGGTATGATCATAAGGAAC
 CGGTGGGCTCGAAGCAGAGGTGCGAGCAGAAGGCGACAAACATAA

>G1524 Amino Acid Sequence (conserved domain in AA coordinates:49-110)

MGRTEQATLTRYPPCPNPAKFNDINKALQEKGYGKALKRKPWTGVTCPVCLEVPHNSV
 VLLCSSYHKGCRPYMCATGNRFSNLCLEQYKKAYAKDEKSDKPELLCPLCRGQVKGWTVV
 EKERKYLNSKKRSCMNDECLFYGSYRQLKKHVKENHPRAKPRAIDPVLEAKWKLEVERE
 RSDVISTVMSSTPGAMVFGDYVIEPYNGYDHQDDSDSDSSDDEMBGGVFELGAFDLGR
 LQPRSAAISSRGIRGMIIRNRWARSRGASRRRQT*

>G1243 (1..3174)

ATGGCGAGAAATTCGAATTCGGATGAGGCTTTCTCGTCAGAGGAGGAAGAAGAGCGGGTT
 AAGGATAATGAAGAAGAAGATGAGGAGGAGCTCGAGGCTGTTGCTCGTCTTCTGGCTCC
 GACGATGACGAAGTAGCCGCCGCCGACGAATCACCAGTCTCCGACGAGAGGCTGCTCCC

GTAGAAGATGATTACGAGGACGAAGAAGATGAGGAAAAAGCTGAAATCAGCAAACGTGAG
AAAGCCAGACTTAAAGAGATGCAGAAGTTGAAGAAGCAGAAGATTCAAGAGATGCTGGAG
TCGCAGAATGCTTCCATTGACGCGGATATGAACAATAAGGGAAAAGGGAGACTGAAGTAT
CTTCTGCAGCAAACCTGAGTTATTTGCCCACTTTGCTAAAAGTGATGGATCTTCTTCTCAG
AAGAAGGCAAAAGGAAGGGGACGTCATGCTTCCAAAATAACTGAAGAGGAGGAAGACGAA
GAGTATCTAAAGGAAGAAGAGGATGGCTTAACTGGATCTGGAACACACGGTTACTCACA
CAGCCCTCTTGATTCAAGGGAAGATGAGAGATTACCAATTAGCTGGTTTGAAGTGGCTC
ATTCGTCTTTATGAGAATGGCATAAATGGAATTCTTGCTGATGAAATGGGTCTGGGGAAG
ACGCTTCAAACGATTTCTTTGTTGGCATATCTTCATGAATACAGGGGAATCAATGGTCCC
CATATGGTGGTTGCTCCAAAATCAACACTTGGTAATTGGATGAACGAAATTCGCCGTTT
TGTCCTGTCTACGTGCTGTGAAGTTCCTTGGTAATCCTGAGGAGAGGAGACATATTCGA
GAAGACCTGCTAGTTGCTGGGAAATTTGATATTTGTGTCACAAGCTTTGAGATGGCCATC
AAAGAGAAGACAGCACTTCGTCGGTTTAGCTGGCGTTATATTATCATTGATGAAGCGCAT
CGAATCAAGAACGAGAATTCACCTCTTCTAAAACCATGAGACTTTTTCAGCACAATTAT
CGGCTTCTTATCACGGGGACCCCTTCAGAATAATCTCCATGAAGTGTGGGCTCTTCTA
AATTTTCTTCTGCTGAGATTTTGTGTCAGCAGAGACTTTTGATGAATGGTTTCAAATT
TCTGGTGAGAATGACCAGCAAGAAGTTGTGCAACAACCTGCACAAGGTTCTTCGACCATT
CTTCTTCAAGACTAAAGTCAGATGTTGAGAAAGGTTTGCACCGAAGAAGGAGACCATA
CTTAAAGTTGGTATGTCTCAGATGCAAAAGCAATACTACAAGGCTTTACTGCAGAAGGAT
CTTGAAGCGGTTAATGCTGGTGGAGAACGCAACGCTCTGCTAAACATTGCAATGCAACTG
CGTAAATGCTGCAATCACCCCTATCTCTTCCAGGTGTCAGAACCTGGTCCCCCATATACC
ACAGGAGATCACCTTATAACAAATGCTGGTAAGATGGTTCTCTTGGATAAATTGCTTCT
AAGTTGAAAGAACGTGATTCAAGGGTGTCTGATATTTTCTCAGATGACAAGACTTTTGGAT
ATTCTTGAGGACTATTTAATGTATCGTGGTTACTTGTATTGCCGTATTGATGGAACACT
GGTGGTGACGAACGAGATGCCTCCATAGAAGCCTACAACAAGCCAGGAAGTGAGAAATTT
GTTTTCTTGTATCTACTAGAGCTGGAGGGCTTGGTATCAATCTTGCTACTGCAGATGTT
GTGATCCTTTACGATAGTGATTGGAACCCACAAGTCGACTTGCAAGCTCAGGATCGTGCC
CATAGGATTGGTCAAAAAAAGAAGTTCAAGTGTTTCGATCTGCACTGAGTCTGCTATT
GAGGAGAAAGTGATTGAAAGAGCTTACAAGAAGTTAGCACTTGATGCTCTGGTTATTCAA
CAAGGGAGATTGGCAGAACAGAAAAGTAAGTCTGTCAATAAGGATGAGTTGCTTCAAATG
GTAAGATATGGTGTCTGATGGTGTTCAGTTCTAAAGATAGCACAATCACAGACGAGGAT
ATTGATAGAAATCATTGCCAAAGGAGAGAGGCAACAGCTGAACCTTGATGCTAAGATGAAG
AAATTCACAGAAGATGCTATACAGTTTAAATGGATGACAGTGCTGACTTCTATGATTTT
GATGATGACAATAAGGATGAAAACAAGCTCGATTTTAAAAAGATTGTAAGCGACAATTGG
AATGATCCCCCAAGCGGGAGAGAAAGCGCAACTACTCTGAATCTGAGTACTTTAAGCAA
ACATTGCGGCAAGGTGCTCCAGCTAAACCTAAAGAGCCTAGAATTCCGCGCATGCCCCAG
TTGCACGATTTCCAGTTCTTTAACATTGAGAGATTGACCGAGTTGTATGAAAAGGAAGTA
CGTTATCTCATGCAACACATCAGAAAAATCAGTTGAAAGACACAATTGATGTTGAAGAA
CCAGAAGTTGGGATCCCTTAACACTGAAGAAGTAGAAGAAAAGGAGGGATTATTGGAG
GAGGGTTTCTCAACATGGAGCAGAAAGAGATTTTAATACTTTCTCAGGGCTTGTGAGAAG
TATGGCCGCAACGACATAAAAAGCATTGCCTCTGAGATGGAAGGGAAAACAGAGGAAGAA
GTTGAAAGATATGCCAAAGTATTTAAAGAGCGGTACAAGGAGCTGAACGACTATGATAGA
ATCATTAAAGACATTGAGAGGGGAGAGGCAAGGATCTCTAGGAAAGACGAAATCATGAAG
GCCATAGGGAAGAACTGGATCGCTACAGAAACCTTGGCTGGAAGTGAAGATTCAATAT
GGTCAGAACAAAGGCAAGCTGTACAATGAAGAGTGTGACCGTTTCATGATCTGCATGATT
CACAACTTGGTTATGGGAATTGGGATGAGCTAAAGGCAGCATTTAGGACATCGTCTGTG
TTCAGTTTGAAGTGGTTTGTGAAATCCCGCACGAGTCAGGAACCTTGCAAGAAGATGCGAC
ACTCTGATTGCTGATGACGAGAAAGAGAACAGGAGTTTGATGAAAGAGAGAGGCAAGCC
CGCAAAGAGAAGAAGCTCGCGAAGAGTGCAACACCATCAAAGCGACCTTTAGGAAGACAA
GCAAGTGAGAGTCTTTCATCGACGAAGAAGCGGAAGCACCTGTGATGAGATGA

>G1243 Amino Acid Sequence (domain in AA coordinates: 216-609)
MARNSNSDEAFSSEEEERVKDNEEEDDEEELEAVARSSGSDDEVAADSPVSDGEAAP
VEDDYEDEEDEEKAEISKREKARKEMQKLKKQKIQEMLESQNASIDADMNNKGKGRLLKY
LLQQTLELFAHFAKSDGSSSQKAKGRGRHASKITEEEDEEYLKEEEDGLTGSGNTRLLT
QPSCIQKMRDYQLAGLNWLIRLYENGINGILADEMGLKTLQTISSLAYLHEYRGINGP
HMOVAPKSTLGNWMEIRRFVPLRAVKFLGNPEERRHIREDLLVAGKFDICVTSFEMAI
KEKTALRRFSWRYIIIDEAHRIKNENSLSKTMRLFSTNYRLLITGTPLQNNLHELWALL

NFLLPEIFSSAETFDEWFQISGENDQQEVVQQLHKVLRPFLRLRLKSDVEKGLPPKKETI
 LKVGMSQMOKQYYKALLQKDLAVNAGGERKRLNLNIAMQLRKCCNHPYLFQGAEPGPPYT
 TGDHLITNAGKMLLDKLLPKLKERDSRVLIFSQMTRLDDILEDYLMYRGYLYCRIDGNT
 GGDERDASIEAYNKPGESEKFVLLSTRAGGLGINLATADVILYSDWNPQVDLQAQDRA
 HRIGQKKEVQVFRFCATESAIEKVIERAYKKLALDALVIQQRLAEQKSKSVNKDELLQM
 VRYGAEMVFSKSDSTITDEIDIRIIAKGEEATAELDAKMKKFTEDAIQFKMDDSADFYDF
 DDDNKDENKLDFFKIVSDNWNPPKRRERKRNYSSEYFKQTLRQGAPAKPKPRIPRMPQ
 LHDFQFFNIQRLTELYEKEVRYLMQTHQKNQLKDTIDVEEPEGGDPLTTEEVEEKEGLLE
 EGFSTWSRRDFNTFLRACEKYGRNDIKSIASEMEGKTEEEVERYAKVFKERYKELNDYDR
 IIKNIERGEARISRKDEIMKAIGKKLDYRNPWLELKIYQGNKGKLYNEECDRFMICMI
 HKLGYGNWDELKAAFRSTSSVFRFDWFKSRTSQELARRCDTLIRLIEKENQEFDERERQA
 RKEKKLAKSATPSKRPLGRQASESPSSSTKKRKHLSMR*

>G631 (190..1461)

CTTCTCTCTTCTTCTTCTTCTTCTTCTCTCTCTCGTCGGATCTCTCTGATTAGTG
 ATTTTTCAAATTTCAAGTTTCTTACCTTTAATTTTGTGTCTCGTTGATCTCTCTTTGG
 ACATCTGCTTTGGATTCTGGAGGCTTCTCATTAGATCTCTATTAGTGGGTTTAGGTCAA
 GTTCTTGAAATGGATAAGGAGAAATCTCTGCACCACCACCTAGTGGAGGTCTTCTCTCA
 CCATCGGGTCGTTACTCTGCGTTTTCACCTAATGGAAGTAGCTTTGCAATGAAAGCTGAA
 TCATCTTTTCTCTCTTCTGACTCCAAGTGGAAGCAATAGCTCAGATGCTAACCGATTGAGC
 CATGATATTAGCCGAATGCCGGATAATCCACCTAAGAACCTAGGCCATCGCCGAGCTCAT
 TCAGAGATTCTTACTCTTCTCTGATGACTTAAGCTTTGATAGTGATCTTGGTGTGGTTGGT
 GCTGCTGATGGACCTTCTTCTCTGATGATACTGACGAGGACTTACTCTATATGTATCTT
 GATATGGAAAAATTCAATTCTTCTGCTACATCGACTTCTCAAATGGGTGAGCCATCAGAA
 CCGACTTGGAGGAATGAATTAGCCTCGACTTCTAACCTTCAGAGTACACCCGGTAGCTCT
 AGTGAAAGACCGAGAATTAGACACCAACACAGCCAATCGATGGATGGTTCAACAACATATC
 AAGCCTGAGATGCTTATGTGACGGAATGAAGATGTGTCTGGAGTTGACTCTAAGAAAGCC
 ATCTCTGCTGCTAAACTTTCTGAGCTTGTCTCTCATTGATCCAAAACGCGCCAAGAGGATA
 TGGGCAAAACAGGCAGTCTGCTCGGAGGTCAAAGAAAGGAAGATGAGATACATTGCAGAG
 CTCGAGAGAAAAGTACAGACTTTACAAACAGAGGCCACATCTCTCTCAGCCCGATTGACT
 CTCTTACAGAGAGATACAAATGGCCTGGGTGTTGAAAACAATGAGCTTAACTGCGAGTA
 CAAACTATGGAGCAACAGGTCACCTACAGGATGCTTAAATGATGCACTAAAGGAGGAA
 GTCCAGCATCTTAAGGTATTGACGGGGCAAGGTCCATCAAATGGTACATCAATGAACCTAC
 GGTCTCTTTGGATCAAACCAGCAATTCTATCCCAATAATCAGTCGATGCACACTATCTTA
 GCCGCACAACAGTTACAGCAGCTCCAGATCCAGTCACAGAAACAGCAACAACAACAACAG
 CAACACCAGCAACAACAACAGCAGCAGCAGCAATTTCACTTTCAACAGCAGCAACTG
 TACCAGCTTCAGCAGCAGCAACGCTTCAACAACAGGAACAACAAGCGGGGCTTCAGAG
 CTAAGAAGACCCATGCCTTCTCTCTGGTCAGAAAGAGAGTGTGACATCGCCTGATCGTGAA
 ACTCCCTTGACAAAAGACTGAGTCTAGACTGTGCTAATGTCCAATTTAGTAAGTTACTCT
 TGGAAAATCTTCTTTTTCATCGCAGGCTCATGGATTGGGATTACTGCATTATAGAGTT
 AAAAACAAGACAGCTTAGAAGTTGCGGATTAGAAAGTTGTTAGTGAAGCTTTTGTCTCG
 TCTGTTGGTAGTTTACAATCTTCTCTTTGTATGATCCTAAG

>G631 Amino Acid Sequence (domain in AA coordinates: TBD)

MDKEKSPAPPPSGGLPPPSGRYSAFSPNGSSSFAMKAESSFPPLTPSGSNSSDANRFSHDI
 SRMPDNPPKNLGHRRHSEILTLPPDLSFSDSLGVVGAADGPSFSDDTDEDLLMYLDME
 KFNSSATSTSQMGEPEPTWRNELASTSNLQSTPGSSSERPRIHQHSQSMGSTTIKPE
 MLMSGNEDVSGVDSKKAISAAKLSLALIDPKRAKRIWANRQSAARSKERKMRYIAELER
 KVQTLQTEATSLSAQLTLLQDRTNGLGVENNELKLRVQTMEEQVHLQDALNDALKEEVQH
 LKVLTLGQGPSNGTSMNYGSFSGSNQQFYFNNQSMHTILAAQQLQQLQIQSQKQQQQQQHQ
 QQQQQQQQFHFQQQQLYQLQQQRLQQQEQQSGASELRRPMPSPGQKESVTSPPDRETPL
 TKD*

>G1909 (1..828)

ATGGGTGGATCGATGGCGGAGAGAGCAAGGCAGGCCAACATTCTCCACTAGCGGGACCC
 CTAAAGTGTCTCGATGCGACTCCAGCAACACTAAGTTCTGTTACTACAACACTATAAC
 CTCCTCAGCCTCGTCACTTCTGCAAAGGTTGCCGTCGCTACTGGACACAAGGGGGCGCC
 CTGAGAAACGTCCCTGTAGGTGGAGGCTGCCGGAGGAATAACAAGAAGGGCAAAAATGGA
 AATTTAAATCTTCTTCTTCTCGTCCAAACAGTCTTCTCTCGGTCAACGCTCAAAGTCCT
 AGCTCAGGACAGCTAAGGACAAATCATCAGTTCCCTTTTCACTCACTCTTTACAATCTC

ACTCAACTCGGAGGTATTGGTTTGAACCTAGCCGCTACTAATGGCAACAACCAAGCTCAC
 CAGATCGGTTCCAGTTTGATGATGAGCGATCTAGGGTTTCTCCATGGACGAAATACTTCA
 ACTCCGATGACGGGAAACATTTCATGAAAAACAACAATAATAACAATGAAAAACAACCTA
 ATGGCATCCGTTGGATCTTTGAGCCCTTTGCTCTCTTCGATCCAACGACGGGGCTATAC
 GCTTTCCAGAACGACGGTAATATCGGGAACAACGTTGGGATATCTGGTTCTTCTACTTCC
 ATGGTTGATTCTAGGGTTTATCAGACGCCTCCGGTGAAGATGGAAGAACAACCTAATTG
 GCTAACTTGTCTAGACCGGTCTCCGGTTTGACGTCTCTGGGAATCAAACAATCAGTAC
 TTTTGGCCTGGTTCCGATTTCTCGGGTCTTCTAATGATCTCTTGTGA

>G1909 Amino Acid Sequence (conserved domain in AA coordinates: 23-51)

MGGSMAERARQANIPPLAGPLKCPKCDSSNTKFCYNNYNLTQPRHFCKGCRRYWTQGGG
 LRNVVPVGGGCRNRNKKGNLKLSSSSSSKQSSSVNAQSPSSGQLRTHQFPFSPTLYNL
 TQLGGIGLNLAAATNGNNQAHQIGSSLMSDLGFLHGRNTSTPMTGNIHENNNNNNNENNL
 MASVGSLSPPALFDPTTGLYAFQNDGNIGNNVGISGSSSTMVDSRVYQTPPVKMEEQPNL
 ANLSRPVSGLTSPGNQTNQYFWPGSDFSGPSNDLL*

>G1663 (64..630)

TTCTCTCTGTGAATCCTTGTTTCATCGTCACTGAAATTAGTTTACAAAATCGACGAATTCG
 GAGATGATTTTTTCAGAATGTGTGCAGAAATGAGTCCAACCTCAACGCTATAGCTTCCGAA
 TCGCGTTCCCAAACGCAGTTCCGTGTTTCGAAATCCTCCTCGAGCGGCGCGGATGTATC
 TCCGCCAGGACTAAAGACCGTCACACGAAGGTTAACGGACGAAGCCGTCGAGTTACGATG
 CCGGCTCTCGCCGCGCTAGGATTTTCCAGTTAACCGGTGAGCTCGGTACAAAACCTGAA
 GGAGAAACCATCGAATGGCTTCTTAGTCAAGCTGAACCGTCGATTATTGCCGCCACTGGC
 TACGGGACTAAGCTCATTTTCGAATTGGGTTGATGTTGCGGCGGACGATTCTCTCGTCGTCG
 TCGTCGATGACGTGCGCGCAAACGCAAACGCAAACGCCACAATCGCCGAGTTGTAGGTTG
 GATCTTTGTGACCAATCGGAATTCAGTATCCGGTGAATGGTTACAGTCATATGCCGTTT
 ACAGCGATGCTTTTAGAGCCGATGACCACGACGGCGGAATCTGAGGTTGAGATCGCGGAG
 GAGGAGGAACGTAGACGCCGTCACCATTAGTAAAATTAGGCTTTTGATTTAGAGTGTTAA
 AATTAGGATTTTAAAGTTTAGGAGGTAACAGATAAGGATAATT

>G1663 Amino Acid Sequence (domain in AA coordinates: TBD)

MIFQNVCRNESNFNAIASESRSQTFGVSKSSSSGGGCSISARTKDRHTKVNGRSRRVTMP
 ALAAARIFQLTRELGHKTEGIEWLLSQAEPSIIAATGYGTLKISNVWDVAADSSSSSS
 SMTSPQTQTPTQSPSCRDLCPIGIQYPVNGYSHMPFTAMLLPEMTTAESEVEIAEE
 EERRRRHH*

>G1231 (103..870)

CAAAACCAAATCTCTCAGCGCCGGTCAAATACTTGTCTCTCTCTCTCTCTCTTTTAC
 TCTTGTCTTGTCTCCTTCGAAGCTGTTTGTCTGTGAAGAAAGATGGAAGCAGGTGGCGCG
 TACAATCCACGCACTGTTGAAGAGGTGTTTAGGGATTTTAAAGGTCGTAGAGCTGGCATG
 ATTAAGGCTTTAACCCTGATGTTTCCAGAGTTTTCGACTTTGTGATCCCGAAAAGGAG
 AACCTTTGCTTTTACCGACATCCAAATGAGCACTGGGAAGTGAATTTGCCAGCTGAAGAG
 GTTCTCTCTGAGCTCCAGAGCCTGTCTTGGGTATCAATTTGCCAGAGACGGGATGGCG
 GAAAAGGATTGGTTGTCCCTTGTGTGCTGCCACAGTGATGCTTGGCTTCTGTGCTGTGCT
 TTCTTTTTTGGAGCCAGGTTTGGATTTGACAAAGCTGATAGGAAGAGGCTTTTCAATATG
 GTGAATGACCTCCCAACAATCTTTGAGGTTGTAGCTGGCACTGCTAAGAAACAAGGAAAA
 GATAAGTCCTCTGTTTCCAACAACAGCAGCAACAGATCCAAATCAAGCTCCAAGCGAGGA
 TCTGAATCCCGTGCCAAGTTCTCAAAGCCGGAGCCCAAGATGATGAGGAGGAGGAAGAG
 GAAGGTGTGGAAGAGGAGGATGAGGATGAGCAAGGTGAAACACAGTGTGAGCATGTGGT
 GAGAGCTATGCAGCTGATGAGTTCTGGATTGCTGTGACCTCTGTGAGATGTGGTTTCAT
 GGAAAGTGTGTTAAGATAACACCAGCAAGAGCTGAGCACATCAAGCAATACAAGTGCCCT
 TCTTGACAGCAACAAAAGGGCTCGTTCCTAAATTTGTTGACCGCTCGTCTGTGTATCTA
 CCTTTGCATATGATGATGAACAGCTTAAGTGTGTTGTTAGATCAGATTGTTCATATGGA
 TTTGGTAATTTAGGAAGACATTTTAGTTTTTTTCATTGTTACATTTTGGCGATTGAAGGGA
 TAACTCTTTGTTTAGGGGTAATGATCTTTTGTCTGTTTATGTTTGTATTAAACATT
 TTCAAACCTCAATCAAAGTATTTTGGTTAGTCTTAAAA

>G1231 Amino Acid Sequence (domain in AA coordinates: TBD)

MEAGGAYNPRTVBEVFRDFKGRRRAGMIKALTTDVQEFFRLCDPEKENLCLYGHNPNEHWEV
 NLPAAEVVPELPEPVLGINFARDGMAEKDWLSLVAVHSDAWLLAVAFFGARFGFDKADR
 KRLFMVNDLPTIFEVVAGTAKKQKDKSSSVSNSSNRKSSSKRGSESRAKFSKPEPKD
 DEEEEEEGVEEEDDEQGETQCGACGESYAADEFWICCDLCEMWFHGKCVKITPARAEHI

KQYKCPSCSNKRARS*

>G227 (21..983)

GTACCGTCGACGATCCGGCGATGTCAAACCCGACCCGTAAGAATATGGAGAGGATTAAAG
GTCCATGGAGTCCAGAAGAAGATGATCTGTTGCAGAGGCTTGTTCAGAAACATGGTCCGA
GGAAGTGGTCTTTGATTAGCAAATCAATCCCTGGACGTTCCGGCAAATCTTGTCTCTCC
GGTGGTGTAAACCAGCTATCTCCGGAGGTAGAGCACCCTGCTTTTTCGCAGGAAGAAGACG
AGACGATTATTTCGAGCTCACGCTCGGTTTGGTAACAAGTGGGCTACGATCTCTCGTCTTC
TCAATGGACGAACCGATAACGCTATCAAGAATCATTGGAACCTCGACGCTGAAGCGAAAAT
GCAGCGTCAAGGGCAAAGTTGTGATTTTGGTGGTAATGGAGGGTATGATGGTAATTTAG
GAGAAGAGCAACCGTTGAAACGTACGGCGAGTGGTGGTGGTGTCTCGACTGGCTTGT
ATATGAGTCCCGGAAGTCCATCGGGATCTGACGTCAGCGAGCAATCTAGTGGTGGTGCAC
ACGTGTTTAAACCAACCGTTAGATCTGAGGTTACAGCGTCATCGTCTGGTGAAGATCCTC
CAACTTATCTTAGTTTGTCTCTTCTTGGACTGACGAGACGGTTCGAGTCAACGAGCCGG
TTCAACTTAACCAGAATACGGTTATGGACGGTGGTTATACGGCGGAGCTGTTTCCGGTTA
GAAAGGAAGAGCAAGTGGAGTAGAAGAAGAAGCAAGGGGATATCTGGTGGATTTCG
GTGGTGAATTCATGACGGTGGTTCAGGAGATGATAAGGACGGAGGTGAGGAGTTACATGG
CGGATTTACAGCGAGGAAACGTCGGTGGTAGTAGTTCTGGCGGCGGAGGTGGCGGTTTCGT
GTATGCCACAAAGTGTAACAGCCGTCGTGTTGGGTTTAGAGAGTTTATAGTGAACCAA
TCGGAATTGGGAAGATGGAGTAGGCGGCC

>G227 Amino Acid Sequence (domain in AA coordinates: 13-112)

MSNPTRKNMERIKGPWSPEEDDLLQRLVQKHGPRNWSLISKSIIPGRSGKSCRLRWCNQLS
PEVEHRAFSQEEDETIIRAHARFGNKWATISRLNGRDINAIGNHWNSTLKRKCSVEGQS
CDFGGNGGYDGNLGEELKRTASGGGGVSTGLYMSPGSPSGSDVSEQSSGGAHVFKPTV
RSEVTASSSGEDPPTYLSLSLPWTDVTVRVNEPVLNQNVTMDGGYTAEFPVRKEEQVE
VEEEEAKGISGGFGGEFMTVVQEMIRTEVRSYMADLQRGNVGGSSSGGGGGGSCMPQSVN
SRRVGFREFIVNQIGIGKME*

>G1842 (219..809)

ACTATTACATGCCTCTTCTCGCTTCAAACCGGCACCGTTTCCACTTGTATTATTTTTTC
TCTCTATCGTCTAACAAAAAAGCTGACTTGGGATTTTTTTTCAATTGTCTAGCCCA
AAAGAAAGATAGAAACGAAGAAAAAAGCAAACACATTTTGGGTCCCCGGTGGTTAGG
ATCAAATTAGGGCACAAACCTTATCGGAGAAAGAACCATGGGAAGAAGAAAAGTCGAGA
TCAAGCGAATCGAGAAACAAAGCAGTCGACAAGTCACTTTCTCAAACGACGCAAAGGTC
TCATCGAAAAAGCTCGACAACCTTCAATTCTCTGTGAATCTTCCATCGCTGTTGTGCGCCG
TCTCCGTTCCCGGAAAACTCTACGACTCTGCCTCCGGTGACAACATGTCAAAGATCATTG
ATCGTTATGAAATACATCATGCTGATGAACCTTAAAGCCTTAGATCTTGCAGAAAAAATTC
GGAATTATCTTCCACACAAGGAGTTACTAGAAATAGTCAAAGCAAGCTTGAAGAATCAA
ATGTCGATAATGTAAGTGTAGATTCTCTAATATCTATGGAGGAACAGCTCGAGACTGCTC
TGTCACTAATTAGAGCTAAGAAGACAGAACTAATGATGGAGGATATGAAGTCACTTCAAG
AAAGGGAGAAGTTGCTGATAGAAGAGAACAGATTCTGGCTAGCCAGGTGGGGAAGAAGA
CGTTTCTGGTTATAGAAGGTGACAGAGGAATGTACGGGAAAATGGCTCCGGCAACAAAG
TACCGGAGACTCTTTCGCTGCTCAAGTAATCACCATCATCAACGGCTGAGCTTTCACCAT
AAACTTACTCACAGCTGATTTCAGAAGCTTTTACAAAATTGTAAATTATAAAAAGCTGCA
TAATAATCTCAACCTTTTTATCTTCTCTCGCCCAATGTGGAAATAAAGGTAAAAACAAAC
GAAGCTCTTTCTTTTATGCGAAAGAATTGTAAACTAAGATAAAGCTACCGATCTTTGT
TGTACCTTAGTAGACAAATATCAGAGTTCTTGTGCTTGT

>G1842 Amino Acid Sequence (domain in AA coordinates: 2-57)

MGRRKVEIKRIENKSSRQVTFSKRRKGLIEKARQLSILCESSIAVVAVSGSGKLYDSASG
DNMSKIIDRYEIHHADELKALDLAEKIRNYLPHKELLEIVQSKLEESNVDNVSVDLSISM
EEQLETALSVIRAKKTELMEDMKSLQEREKLLIENQILASQVGKKTFLVIEGDRGMSR
ENGSGNKVPETLSLLK*

>G1505 (1..681)

ATGGATGATATAGCGGAACCTGAATGGTTATCAAATTCGTAGATGATTCTTCTTTCACG
CCGTATTCTGCTCCGACGAATAAACCGGTTTGGTTAACCGGAAATCGGAGACATCTTGTA
CAACCGGTTAAAGAGGAGACCTGCTTCAAATCCCAACATCCGGCCGTCAAACCCAGACCC
AAACGAGCCAGAACCGGAGTCAGAGTCTGGTCTCATGGTTCCGAGTCGTTAACCGACTCA
TCTTCAAGCTCTACAACATCTTCTGCTGCTCTCTCTCGTCTTCAAGCCCTCTATGGCTC
GCCAGCGGTGAGTTTCTTGATGAGCCAATGACTAAACACAAAAGAAGAAGAAAGTTTGG

AAAAACGCTGGTCAGACGCAAACGCAAACGACGACGACGCGGCAGTGTGGTCATTGT
GGAGTTTCAGAAAACGCCCGCAGTGGAGAGCAGGACCATTAGGAGCGAAGACGTTGTGTAAAT
GCGTGTGGTGTGCGTTACAAATCGGGTCGGTTACTACCCGAATATAGACCCGCTTGTAGC
CCAACATTTTCGAGTGAGCTTCACTCAAACCACCACAGTAAAGTCATTGAGATGCGTAGG
AAGAAAGAGACTTCTGACGGTGCTGAAGAAACCGGTTTGAACCAGCCGGTTCAGACGGTT
CAGGTTGTCTCGAGTTTTTGA

>G1505 Amino Acid Sequence (domain in AA coordinates: TBD)
MDDIALEWLSNFVDDSSFTPYSAPTNKPVWLGNRRHLVQPVKEETCFKSQHPAVKTRP
KRARTGVRVWSHGSQSLDSSSSSTSSSSSPRPSPLWLASGQFLDEPMTKTQKKKVV
KNAGQTQTQTQTQTRQCGHCGVQKTPQWRAGPLGAKTLCNACGVRYKSGRLLPEYRPACS
PTFSSELHSHNSKVIEMRRKETS DGA EETGLNQPVQTVQVVSSF*

>G657 (1..2331)
ATGAAGCGTGAGATGAAAGCACCTACTACTCCACTAGAGAGTCTCCAAGGTGACCTCAAA
GGAAAACAAGGGAGGACATCTGGCCCTGCTAGACGATCTACCAAAGGACAATGGACACCT
GAAGAGGACGAAGTCTTGTGTAAAGCTGTTGAGCGTTTTCAAGGAAAGAAGTGGAAAG
ATAGCTGAATGTTTAAAGGATCGGACTGATGTTCAAGTGTCTTCATAGATGGCAAAGGTC
TTGAACCCAGAGCTTGTGAAAGGACCGTGGTCAAAGAGGAGGATAACACAATAATTGAC
CTGGTTGAAAAATATGGGCCAAAGAAATGGTCTACTATATCTCAGCATTACCTGGGCGC
ATAGGAAAGCAATGTAGGAAAGGTGGCATAACCATCTTAACCTGGGATTAATAAAAAAT
GCATGGACTCAGGAAGAGGAACTGACTCTTATTCGTGCGCATCAAATTTATGGGAATAAA
TGGGCAGAGCTTATGAAATTTTGGCAGGAAGGTGAGATAATTCGATAAAAAATCATTGG
AACAGCTCAGTTAAGAAGAAGTTGGATTCTTACTATGCATCAGGTCTTTTAGATCAGTGT
CAAAGCTCGCCATTAAATTTGCCCTTCAGAACAAATCTATCGCTTCATCTTCTCGTGGATG
CACAGCAATGGAGATGAAGGTAGTTCAAGGCCAGGGGTTGATGCTGAGGAATCAGAAATGC
AGCCAAGCTTCAACTGTTTTCTCACAATCAACCAACGATTTACAAGATGAAGTTCAACGT
GGAAATGAGGAATATTACATGCCTGAATTTTCAATCAGGAACGGAGCAGCAAATCTCAAAC
GCTGCATCTCATGCAGAACCGTACTACCTTCTTTAAAGATGTCAAATTTGTTGTCCCC
GAAATTTCTTGTGAAACAGAATGTTTCAAGAAAGTTTCAAGATCTTAATTGTTCTCACGAG
CTAAGAACTACCACAGCTACGGAGGATCAATTGCCGGGTGTATCTAATGATGCTAAACAG
GACCGTGGTCTAGAGTTATTGACCCATAACATGGACAACGGTGGAAAAACCAAGCACTT
CAACAAGATTTTCAAAGTTCAGTAAGATTAAAGTATCAACCTTTTTTGTCAAACCTCGGAC
ACAGATCCAGAAGCTCAAACTTTGATCACGGATGAGGAGTGTGTAGGGTTCTTTTCCA
GATAACATGAAAGATAGCAGTACATCTTCTGGTGAGCAAGGTGCGGAATATGGTTGACCTT
CAAAACGGCAAAGGATCTCTTGTCTCAGGCTGCAGAAACCCATGCTCATGAAACTGGA
AAAGTTCCAGCTTTACCGTGGCATCTTCAAGTTCTGAGGGCCTGGCGGGTCATAATTGT
GTCCCTTTGTTGGATTGAGACTTGAAGGACTCACTTTTACCCGTAATGATTCCAACGCT
CCTATACAAGGTTGTGCGCTTTTGGAGCTACCGAATTAGAATGTAAGACTGATACAAAT
GACGGTTTCATCGATACTTACGGACATGTAACCTTCCATGGCAATGATGATAATGGTGGT
TTCCAGACAACAGGGGCTGTCTATATTTCCCAAGGATTCTTTGAAGCTAGTACCTTTG
AATAGTTTTTCTTCTCCTTCTAGAGTGAACAAGATTATTTTCTTATTGACGATAAGCCG
GCTGAAAAAGACAAAGGAGCTCTTTGTTATGAACCTCCACGTTTTTCCAAGTGCAGATATT
CCTTTCTTCAGCTGTGATCTTGTACCATCAAATAGTGACTTACGGCAAGAGTACAGTCCC
TTTGGTATCCGTGAGTTGATGATTTCTTCAATGAATTGTACAACCTCCGTAAAGGTTATGG
GATTCACCGTGTACGATAGGAGCCCTGATGTCTGCTTAATGATACTGCCAAAAGTTTT
AGTGGTGCACCATCCATCTTAAAGAAGCGGCATCGAGACTTGCTTTACCTGTGCTTGAT
AGAAGAAAAGACAAAAAGCTTAAAGGGCTGCGACTTCTTCTTGGCTAATGATTTTTCG
CGCTTAGATGTAATGCTTGTGATGAAGGAGATGATTGCATGACCTCTCGTCCGTGAGTCT
CCTGAAGATAAAAAATATGATGCTTCCCTTCCATAGCCAGAGATAACAGAAATTGTGCA
TCAGCTCGGTTATATCAAGAAATGATTCCGATAGATGAGGAACCAAGGAAACCTTAGAA
TCAGGTGGAGTGACTTCTATGCAAAATGAAATGGATGTAATGACGGTGGTGTCTCAGCT
AAAAATGTAAGTCCGTCTTTGTCTTGCATATTATCTGGTATCAGTTATAA

>G657 Amino Acid Sequence (domain in AA coordinates: TBD)
MKREMKAPTTPLESLQGLDKGKQGRSTSGPARRSTKGQWTPPEDEVLC KAVERFQKKNWKK
IAECFKDRD TDVQCLHRWQKVLNPELVKGPWSKEEDNTIIDLVEKYGPKKWSTISQHLPGR
IGKQCRERWHNHLNPGINKNAWTQEEELTLIRAHQIYGNKWAELMKFLPGRSDNSIKNHW
NSSVKKLDLSYYASGLLDQCQSSPLIALQNKSIASSSSWMHNSNGDEGSSRPGVDAEES
SQASTVFSQSTNDLQDEVQRGNEEYMPFHSQTEQGISNAASHAEPYYPSPKDVKIVVP

EISCETECSSKKFQNLNCSHELRTTTATEDQLPGVSNDKQDRGLELLTHNMDNGGKNQAL
 QQDFQSSVRLSDQPFLLNSDTPPEAQTLLITDEECCRVLFDPNMKDSSTSSGEQGRNMVDP
 QNGKGSLSQAAETHAHETGKVPALPWHPPSSSEGLAGHNCVPLLDSDLKDSLPRNDSNA
 PIQGCRLFGATELECKTDTNDGFIDTYGHVTSHGNDNGGFPEQQGLSYIPKDSLKLVLPL
 NSFSSPSRVNKIYFPIDDDKPAEKDKGALCYEPPRFPSADIPFFSCDLVPSNSDLRQEYSP
 FGIRQLMISSMNCTTPLRLWDSPOCHDRSPDVMLNDTAKSFSGAPSILKKRHRDLLSPVLD
 RRRKDKLKRAATSSLANDFSRLDVMLDEGDDCMTSRPSESPEKDNICASPSIARDNRNCA
 SARLYQEMIPIDEETKESGGVTSMQENGCNDGGASAKNVSPSLSLHIIWYQL*

>G1959 (141..1028)

CGTCGACTGTCCATAAATCCGGAGCCTGACCCGACGTTTGACCCGGATCCGAAACTCCCA
 CAATCTCCATACCACCCAAATTCATCTCCCTAAAGCTTTCTCTCACTTTCCCGGAAAA
 TCGGCGACCAAAATTGAAAATGTACTCAGCGATTGCTCGCTTCCACTCGATGGTGGAC
 ACGTTGGTGGTGACTACCATGGACCTCTTGACGGAACCAATCTCCCGGTGACGCTTGT
 TGGTTTTAACGACTGACCCTAAACCTCGTCTCCGGTGGACAACCTGAGCTTCATGAGAGAT
 TCGTTGACGCCGTACTCAGCTCGGTGGTCTGACAAAGCGACTCCCAAACTATTATGA
 GAACAATGGGAGTGAAGGGTCTCACTCTTACCACCTCAAATCACATCTTCAGAAATTCC
 GCCTAGGGAGGCAAGCTGGCAAAGAATCAACTGAGAACTCTAAAGATGCTTCTTGTGTAG
 GGGAGAGTCAGGACACAGGTTTCTTTCGACATCATCAATGAGAATGGCGCAGCAGGAGC
 AGAACGAGGGTTACCAAGTCACCGAAGCTCTACGTGCTCAGATGGAAGTCCAAAGAAGAC
 TACACGATCAATTGGAGGTGCAACGGAGGCTCCAGCTGAGGATAGAGGCACAAGGAAAT
 ACCTGCAATCGATTCTTGAAAAGCTTGCAAGGCCTTTGACGAGCAAGCTGCTACTTTTG
 CTGGAATTGAGGCTGCTAGGGAAGAGCTATCAGAGCTAGCCATCAAAGTCTCCAATAGCT
 CTCAGGAACACTCAGTCCCGTACTTCGATGCAACAAAGATGATGATGATGCCATCGTTGT
 CAGAGCTTGCACTAGCAATAGACAACAAAAACAACATCACAACTGTTTCTAGTAGAAA
 GCTCTCTGACTTCCATCACACATGGGAGCTCTATATCTGCTGCATCAATGAAGAAGCGTC
 AACGTGGAGACAATTTGGGCGTAGGGTATGAATCAGGCTGGATTATGCCTAGTAGACCA
 TTGGATAAAGTTTAGGAGAGGGGAAAAAGTTTATATGGGAAAGGTAGAGATAAGATTAA
 CTGTTCTTTACTTGCTTTGAGGGGCTGCGGCCGCT

>G1959 Amino Acid Sequence (conserved domain in AA coordinates:46-97)

MYSAIRSLPLDGGHVGVDYHGLDGTNLPDACLVLTTDPKPRLRWTELHERFVDAVTQ
 LGGPDKATPKTIMRTMGVGLTLYHLKSHLQKFLRGRQAGKESTENSKDASCVGESQDTG
 SSSTSSMRMAQQEQNEGYQVTEALRAQMEVQRRRLHDQLEVQRRQLRIEAQGGYLSILE
 KACKAFDEQAATFAGLEAAREELSELAIKVSNSSQGTSVPYFDATKMMMPSLSELAVAI
 DNKNITTNCSVESSLTSITHGSSISAASMKKRQRQDNLGVGYESGWIMPSSTIG*

>G2180 (1..1440)

ATGGCTCCTGTCTCGTTACCTCCAGGTTTCCGATTCCATCCAACAGACGAGGAACATAATT
 ACTTACTATCTAAAAAGAAAGATCAACGGTCTAGAAATCGAACTTGAAGTTATCGCTGAA
 GTTGATCTTTACAAGTGTGAGCCATGGGACTTACCAGGGAAGTCCTTGCTTCCGAGCAAA
 GACCAAGAATGGTACTTCTTCAGCCCACGAGACCGGAAGTATCCAACGGCTCAAGGACA
 AACCGGGCAACTAAAGGCGGTTATTGGAAGGCTACAGGTAAAGACCGCCGAGTTAGTTGG
 AGAGACCGAGCCATAGGAACCAAGAAGACATTGGTTTACTACCGTGGGCGCGGCCACAT
 GGCATAAGAAGTGGTTGGGTATGCACGAATATCGACTTGATGAAACAGAATGTGAGCCT
 TCTGCATACGGCATGCAGGACGCATATGCACCTTTGTCGTGTGTTCAAAAAGATTGTTATT
 GAAGCTAAGCCAAGAGATCAACATCGGTCATATGTCCACGCGATGTGCAATGTGAGTGGT
 AATTGCTCATCGAGTTTTGACACTTGTTCGGATCTCGAAATCAGTTCAACTACTCATCAA
 GTTCAAAACACATTCCAACCGCGATTGGCAACGAGCGATTAACTCCAACGCAATCAGC
 AACGAGGATTGGTCACAATACTACGGTCTCTTCTATAGACCGTTCCCTACTCCATATAAG
 GTTAACACAGAGATGGAATGTTCAATGTTACAACACAATATATATCTACCACCGTTGCGT
 GTAGAAACTCTGCGTTTAGTGATTCCGATTTCTTCACGAGTATGACTCACAACAACGAC
 CATGGCGTTTTCGATGACTTTTACTTTTGTGCAAGTAACTCCAACCACAATAATAGCGTT
 GGTGATCAAGTGATCCACGTTGGCAATTATGATGAACAATTAATAACATCTAACCGTCAT
 ATGAACCAGACTGGTTATATAAAAGAGCAGAAGATCAGATCGAGTTTGGATAATACTGAC
 GAAGATCCAGGATTTTCATGGTAACAATACCAATGACAACATAGATATCGATGATTTTCTC
 TCGTTTGATATATATAACGAGGACAACGTGAATCAAATAGAAGATAATGAAGACGTGAAT
 ACAATGAACCCCTTGATTTCATCGGGATTTCGAGGTGGTTGAAGAAGAACTAGATTTAAC
 AACCAATGCTCATCTCGACATATCAAACGACAAAGATTCTATATCACCAGTCGTACCT
 TGTACACGTTGAAAGTTACGTCATCTTATTAGTCACAATGTGGAAGAGAGAACATTG

TTCATTGAAGAGGACAAAGATTCTTGGTTACAAAGAGCTGAGAAGATCACGAAGACAAAA
 CTAACACTTTTTAGTTTAATGGCTCAGCAATACTACAAATGTCTTGCTATTTTTTCTGA
 >G2180 Amino Acid Sequence (conserved domain in AA coordinates:7-156)
 MAPVSLPPGFRFHPTDEELITYYLKRKINGLEIELEVIAEVDLYKCEPWDLPKSLPSK
 DQEWYFFSPRDRKYPNGSRITNRATKGGYWKATGKDRRVSWRDRAIGTKKTLVYYRGRAPH
 GIRTGWMHEYRLDETECEPSAYGMQDAYALCRVFKKIVIEAKPRDQHRSVVHAMSNNVSG
 NCSSSFDTCSLEISSTTHQVQNTFQPRFGNERFNSNAISNEDWSQYYGSSYRPFPTPYK
 VNTEIECSMLQHNILPLRVENSAPSDSDFFTSMTHNNDHGVFDDFTFAASNSNHNSV
 GDQVIHVGNYDEQLITSNRHMNTGYIKEQKIRSSLDNTDEDPGFHGNNTNDNIDIDDFL
 SFDIYNEDNVNQIEDNEDVNTNETLDSSGFVVEEETRFNNQMLISTYQTTKILYHQVVP
 CHTLKVHNPISHNVEERTLFIEEDKDSWLQRAEKITKTKLTLFSLMAQQYYKCLAIFF*
 >G1817 (1..1308)

ATGAAGGACGCAGAGAAGCGAGAGGTGATTGCATCATCATCATTACAAAGAAAGAGAAAC
 AGAGGAAGAAGACTAAGGAAAAAGAAAGAAACGAGAAGCGAGTACTAATGGTTCCA
 TCATCATTACCAAACGACGTGCTAGAGGAGATCTTTTAAAGATTTCGGTTAAAGCCCTA
 ATCCGACTCAAGTCTCTCTCGAAACAATGGAGATCGACGATCGAATCTCGCAGTTTTGAA
 GAGAGACACTTGACGATCGCTAAGAAAGCCTTCGTGGATCATCCCAAGGTCATGCTCGTA
 GGAGAAGAAGATCCCATAAGAGGAACCGGGATTTCGTCCAGACACTGACATTGGTTTTAGG
 TTATTCTGCTTGGAATCGGCTTCTCTTCTATCCTTTACTCGTCTCAATTTCCCTCAAGGG
 TTCTTCAACTGGATCTACATATCTGAAAGCTGTGATGGCCTTTTCTGCATCCATTCCCCA
 AAATCACATTCCGTATATGTAGTGAATCCGGCTACACGGTGGCTCCGCCACTTCTCTCCG
 GCAGGGTTTCAGATTTTGATCCACAAGTTTAACCCCACTGAACGTGAGTGAATGTAGTG
 ATGAAATCAATCTTTTCATCTAGCATTCTGTGAAGGCCACCGATTACAAATAGTGTGGTTG
 TACAATTGTGATAAGTACATTGTTGATGCGTCGAGTCCAAACGTGGGAGTCACAAAGTGC
 GAGATTTTGACTTTAGGAAAAATGCTTGAGGTTACTTGGCTTGCACTCCAAGTCATCAG
 ATATTCTATTACCAAAGCCAGCATCTGCAAACGGGTCGGTTTATTGGTTTACAGAACCA
 TATAATGAAAGAATCGAAGTAGTGGCTTTTGATATTCAGACCGAAACATTCCGGTTGCTG
 CCTAAGATTAATCCGGCTATTGCTGGTTTCAGATCCTCACCATATTGACATGTGCACTCTG
 GATAATAGTTTGTGTATGTGCGAAAAGGGAGAAAGATACTATGATCCAAGATATTTGGAGG
 TTGAAACCATCAGAAGACACATGGGAAAAGATTTTAGCATAGACTTGGTTTCTGTCTCT
 TCTTCTCGGACTGAGAAGCGTGATCAATTTGATTGGAGCAAGAAGGATAGGGTTGAGCCA
 GCCACACCCGTCGCGGTTTGTAAGAATAAGAAGATCCTTCTCTCACATCGCTATTCCCGA
 GGTTTGGTAAAGTACGATCCCCTAACAAATCTATCGATTTTTTTTTCCGGACATCCTACC
 GCTTACAGAAAAGTTATTTATTTTCAAAGTTTGATATCTCATCTATAA

>G1817 Amino Acid Sequence (conserved domain in AA coordinates:47-331)
 MKDAEKREVIASSSLQKRNRGRRLRKRNRNKRKVLMPSSLPNDVLEEIFLRFPVKAL
 IRLKSLSKQWRSTIESRSFEERHLTIAKKAFVDHPKVMLVGEEDPIRGTGIRPDIDIGFR
 LFCLESASLLSFTRLNFPQGFNWIYISESCDGLFCIHSPKSHSVYVNPATRWRLRLLPP
 AGFQILLHKFNPTERWNVVMKSIHFLAFVKATDYKLVLVLYNCDKYIVDASSPNVGVTKC
 EIFDFRKNWRYLACTPSHQIFYQKPPASANGSVYWFTEPYNERIEVVAFDIQTETFRLL
 PKINPAIAGSDPHHIDMCTLDNSLCMSKREKDTMIQDIWRLKPSEDTWEKIFSIDLVSCP
 SSRTEKRDQFDWSKKDRVEPATPVAVCKNKKILLSHRYSRGLVKYDPLTKSIDFFSGHPT
 AYRKVIYFQSLISHL*

>G1649 (61..1311)
 ATTCACAAAACCGGAAAAAAGACAAGTAAAGAAAGCTTTGTTTCAGTTTACTTCA
 ATGGAAGCAAAACCTTAGCATCATCATCTGAACCAACATGATTTCTCCATCATCA
 AACATTAACCAAAAATTAAGAGATGAAGATTATATGGAGCTGGTGTGTGAAAATGGGCAG
 ATTCTTGCAAGATTCGAAGACCAAGAACAACGGTTCTTTTCAAAGCAACGTAGGCAA
 TCTCTCCTGGATTTGTATGAGACCGAGTACAGCGAGGGTTTCAAGAAAAACATCAAGATT
 CTTGGAGACACACAAGTTGTTCCGGTGAGTCAGTCTAAGCCACAACAAGATAAAGAAACC
 AATGAACAAATGAACACAATAAGAAGAAGCTAAAGTCTCCAAAATCGAATTTGAGAGA
 AATGTTTCGAAAAGCAACAAATGTGTTGAATCATCAACATTAATTGATGTTTCTGCTAAA
 GGTCCAAAGAATGTTGAAGTTACTACAGCTCCTCCTGATGAGCAATCTGCAGCTGTTGGT
 AGATCCACGGAATTGTATTTGCTTCTTCATCGAAGTTTTCTCGAGGAACCTCGAGAGAT
 CTAAGTTGTTGTTCTTTAAAGAGGAAGTATGGAGATATTGAAGAAGAAGATCAACCTAT
 TTAAGTAATAATTCAGATGATGAATCAGATGATGCGAAGACACAAGTTCATGCGAGAAC
 AGAAAGCCGGTGACTAAAAGAAAACGAAGCACAGAAGTCCATAAGTTATATGAAAGAAAA

CGAAGAGATGAATTCAACAAGAAAAATGCGTGCTTTGCAGGACCTACTACCAAATTGTTAC
AAGGATGATAAGGCTTCATTGTTGGATGAGGCTATCAAATATATGCGGACCCCTCAACTT
CAAGTTCAGATGATGAGTATGGGAAATGGATTAATAAGACCACCTACGATGTTGCCAATG
GGTCATTACTCTCCCATGGGTCTAGGAATGCATATGGGTGCAGCAGCAACACCAACATCA
ATACCGCAATTCTGCTATGAATGTTCAAGCAACCGGTTTCCGGGGATGAACAAATGCA
CCACCACAAATGCTAAGCTTTCTTAATCACCCAAGTGGACTAATTCCAAACACTCCTATC
TTTTCTCCATTGGAAAATTGCTCTCAGCCATTCTGTTGCTTCTGTTGTTTCTCAGACT
CAGGCTACTTCTTTTACTCAATCCCAAAGTCTGCGTCCGCTCAAACCTAGAAGATGCA
ATGCAATATAGAGGAAGCAACGGTTTTAGTTATTATCGCTCGCCAACTAATGATTGTGA
GAAAGTTGATGTTTCTCCAACCTAACTTTAAGCAAAAAAATGATCGTCTACTCT
GTGTTGTTAGTCTATGGGCTTTTGGGCTTGATTCTTGAACGATTTGAACCTAATTCCA
ACTATTTTCAAAGTGGATGTACAAAGTAAAA

>G1649 Amino Acid Sequence (conserved domain in AA coordinates:225-295)

MEAKPLASSSEPNMISPSNNIKPKLKDEDYMLVCENGQILAKIRRPKNNGSFQKQRRQ
SLLDLYETEYSEGFKKNIKILGDTQVVPVSQSKPQDKETNEQMNNNNKKLKSSKIEFER
NVSKSNKCVESSTLIDVSAKGPKNVEVTTAPPDEQSAAVGRSTELYFASSSKFSRGTSRD
LSCCSLKRKYGDIEEEESTYLSNNSDDESDDAKTQVHARTRKPVTKRKRSTEVHKLKERK
RRDEFNKKMRALQDLLPNCYKDDKASLLDEAIKYMRTLQQLVQMMSMGNGLIRPPTMLPM
GHYSMPGLMHMGAAATPTSIPQFLPMNVQATGFGPMNNAAPPQMLSFLNHPGLIPNTPI
FSPLENCSPFVVPSCVSQTQATSFTQFPKSASASNLEDAMQYRGSNGFSYRSPN*

>G2131 (69..1010)

GTCTCTCATTTTCATAATTCATTTTCAGGATTGTCTCTCAATCTTTTATTCTTCTCATT
CACCGGTAAATGGCAAAAGTCTCTGGGAGGAGCAAGAAAACAATCGTTGACGATGAAATCA
GCCATAAAACAGCGTCTGCGTCTGAGTCTGCGTCCATTGCCTTAACATCCAAACGCAAC
GTAAGTCGCGCGCTCGAAACGCTCCTCTTCAACGCAGCTCCCTTACAGAGGCGTCACAA
GGCATAGATGGACTGGGAGATACGAAGCGCATTTGTGGGATAAGAACAGCTGGAACGATA
CACAGACCAAGAAAGGACGTCAAGTTTATCTAGGGGCTTACGACGAAGAAGAGCAGCAG
CAGTGCCTACGACTTAGCAGCATTGAAGTACTGGGACGAGACACACTCTTGAACCTCC
CTTTGCCGAGTTATGACGAAGACGTCAAAGAAATGGAAGGCCAATCCAAGGAAGAGTATA
TTGGATCATTGAGAAGAAAAGTAGTGGATTTTCTCGCGGTGTATCAAAATACAGAGGCG
TTGCAAGGCATCACCATAATGGGAGATGGGAAGCTAGAATTGGAAGGGTGTGTTGGTAATA
AATATCTATATCTTGGAAACATACGCCACGCAAGAAGCAGCAATCGCCTACGACATCG
CGSCAATAGAGTACCGTGGACTTAACGCCGTACCAATTTGACGCTCAGCCGTTATCTAA
ACCCTAACGCCGCCCGCGGATAAAGCCGATTCCGATTCTAAGCCCATTGGAAGCCCTAGTC
GCGAGCCCGAATCGTCGGATGATAACAAATCTCCGAAATCAGAGGAAGTAATCGAACCAT
CTACATCGCCGGAAGTGATTCCAATCGCCGGAGCTTCCCGACGATATCCAGACGTATT
TTGGGTGTCAAGATTCCGGCAAGTTAGCGACTGAGGAAGACGTAATATTGATTGTTTCA
ATTCTTATATAAATCCTGGCTTCTATAACGAGTTTGATTATGGACCTTAATCGTATTTTC
TACAAGTTTTGTTTGGATTATCTACACAATACATCAATATATTCT

>G2131 Amino Acid Sequence (conserved domain in AA coordinates:50-186, 112-183)

MAKVSGRSKTIVDDEISDKTASASESASIALTSKRKRKSPPRNAPLQRSSPYRGVTRHR
WTGRYEAHLWDKNWNTQTQKGRQVYLGAIDEDEEAARAYDLAALKYWGRDILLNFPLP
SYDEDVKEMEGQSKEEYIGSLRRKSSGFSRGVSKYRGVARHHHNGRWEARIGRVFGNKYL
YLGTYATQEEAAIAYDIAAIEYRGLNAVTFNFDVSRYLNPNAADKADSDSKPIRSPSREP
ESSDDNKSPKSEEVIEPSTSPVIPTRRSFPDDIQTYFGCQDSGKLATEEDVIFDCFNSY
INPGFYNEFDYGP*

>G215 (1..1110)

ATGACTCGTCTGGTGTTCGCATTGTAGCAACAATGGGCACAATTCACGCACGTGTCCAACG
CGTGGGTCTGGTTCTCTCCGCGGTGAAGTTATTTGGTGTGAGGTTAACGGATGGCTCG
ATTATTAAGAGATGCGAGTATGGGTAACTCTCTCGGCATTGGCTGTTGCGGCGGCGCG
GCAACGCACCACCGTTTATCTCCGTCTCTCTCTGGCGACGTCAAATCTTAATGATTGCG
CCGTTATCGGATCATGCCCCGATACTCTAATTTGCATCATAATGAAGGGTATTTATCTGAT
GATCTGCTCATGGTCTGGGTCTAGTCACCGTCTGGTGTGAGAGGAAGAGAGGTGTTCTCT
TGGACTGAAGAGGAACATAGACTATTCTTAGTCCGTCTTCAGAACTCGGGAAAGGAGAT
TGGCGCGGTATTTTCGAGAACTATGTAACGTCAAGAACTCTTACACAAGTGGCTAGTCAT
GCTCAAAAGTATTTTATTCGACATACTAGTTCAAGCCGACGAGAAAGACGGTCTAGCCTC
TTCGACATGGTTACAGATGAGATGGTAACCGATTATCGCCAACACAGGAAGAGCAGACC

TTAAACGGTTCCTCTCCAAGCAAGGAACCTGAAAAGAAAAGCTACCTTCCTTCAC'TTGAG
CTCTCACTCAATAATACCACAGAAGCTGAAGAGGTGCTAGCCACGGCGCCACGACAGGAA
AAATCTCAAGAAGCTATAGAACCATCAAATGGTGTTCACCAATGCTAGTCCCGGGTGGC
TTCTTTCCTCCTTGTGTTTTCCAGTGACTTACACGATTTGGCTCCCTGCGTCACTTACGGA
ACAGAACATGCCTTAAACGCTGAGACTTCTTCTCAGCAGCATCAGGTCCTAAACCAAAA
CCTGGATTTGCTAAAGAACGTGTGAACATGGACGAGTTGGTCGGTATGTCTCAGCTTAGC
ATAGGAATGGCGACAAGACACGAAACCGAAACTTCCCTTCCCGCTATCTTTGAGACTA
GAGCCCTCAAGGCCATCAGCGTTTCACTCGAATGGCTCGGTTAATGGTGCAGATTTGAGT
AAAGGCAACAGCGCGATTCAGGCTATCTAA

>G215 Amino Acid Sequence (domain in AA coordinates: TBD)
MTRRCSHCSNNGHNSRTCPTRGSGSSSAVKLFGVRLTDGSIKKASAMGNLSALAVAAAA
ATHHRLSPSSPLATSNLNDSPSLDHARYSNLHHNEGYSDDPAHSGSGSHRRGERKRGVP
WTEEHRLFLVGLQLKLGKGDWRGISRNYVTSRTPTQVASHAQYFIRHTSSSRKRSSL
FDMVTDEMVTDSSTQEEQTLNGSSPSKEPEKKSYPLESLNNTTEAEVAVATAPRQE
KSQEAIEPSNGVSPMLVPGGFFPPCFPVYTIWLPASLHGTEHALNAETSSQQHQVLKPK
PGFAKERVNMDLVGMSQLSIGMATRHETETSPSPLSLRLEPSRPSAFHSNGSVNGADLS
KGNIAIQAI*

>G1508 (1..420)

ATGCTAGATCACAGTGAAAAGGTCTTATTGGTTGATTAGAAAACCATGAAAACAAGAGCT
GAAGATATGATCGAACAGAACAACTAGTGTAAACGACAAGAAGAAGACTTGTGCTGAT
TGTGGAACCAAGTAAACTCCTCTTTGGCGTGGTGGTCTGTGGTCCAAAGTCGTTGTGT
AACCGGTGTGGGATCAGAAACAGAAAGAGAAGAGGAGGAACAGAGATAATAAGAAA
TTAAAGAAATCGAGTTCTGGCGGCGGAAACCGTAAATTTGGTGAATCGTTAAACAGAGT
TTGATGGATTTGGGGATAAGGAAGAGATCAACGGTGGAGAAGCAACGACAGAAGCTTGGT
GAAGAAGAACAAAGCCGTGTGTTACTCATGGCTCTTTCTTATGGCTCTGTTTACGCTTAG

>G1508 Amino Acid Sequence (domain in AA coordinates: 38-63)
MLDHSEKVLVLDSETMKTRAEDMIEQNNTSVNDKKKTCADCGTSKTPLRGGPVGPKSLC
NACGIRNRKRRGGTEDNKKLKKSSSGGNRKFESLQKSLMDLGIRKRSTVEKQRQKLG
BEEQAAVLLMALSYGSVYA*

>G2110 (36..1622)

GAGAGCTAATAAAAAATTTATCAAAGAAGACTAATATGGAGAAGGACGATTTCTTGAGGA
GTGGTCATGGAAGAGAAGAAAGCCATGATGAGATGAGAAAACCTTGATTCATCTCAGCATG
ATTCTCATCAAGAACACGACCATATTATAAGATCCAAGTTGGACTCAACTAAAGTCGAAA
TGGATGAGGCTAAAGAGGAAAATCGAAGACTAAAGTCATCATTGAGTAAATCAAGAAAG
ATTTTGACATCCTTCAAACACAATACAACCAATTAATGGCCAAACATAACGAACCAACCA
AGTTCCAATCAAAAGGGCATCATCAAGACAAAGGCGAAGATGAAGACAGAGAAAAAGTTA
ACGAACGTGAAGAACTTGTCTCGTTGAGCCTAGGCAGACGGTTAAATTGAGAGGTTCCAA
GTGGTTGGAATAAAGAAAGCAAAATAAAGATGTTGAAGAAGCGGAAGGTGACAGAAATT
ATGATGATAATGAAAAAAGCAGTATTCAAGGGTTGAGTATGGGGATTGAATACAAGGCTT
TGAGTAATCCTAATGAGAAGTTAGAGATTGATCATAATCAAGAAACCATGTCGTTGGAGA
TTAGTAACAATAATAAGATCAGATCAGAAAATAGTTTGGGTTTAAAGATGATGGAGATG
ATCATGAAGATGAAGATGAGATTTGCCTCAAACCTTGTTAAGAAAACCTAGGGTTTCGG
TGAGATCAAGATGTGAGACACCAACGATGAACGACGGATGTCAATGGAGGAAATATGGCC
AGAAAATAGCTAAAGGCAATCCATGTCCCCGAGCTTACTATCGTTGCACCATGCGAGCTT
CTTGTCCAGTAAGAAAACAGGTGCAAAGATGTTTCAAGATATGTCTATACTTATCTCAA
CGTACGAAGGAACACATAACCATCCACTTCCCATGTCAGCAACTGCCATGGCCTCTGCCA
CTTCCGCTGCCGCTCCATGCTTCTCTCCGGCGCCTCCTCCTCATCCGCCGAGCTG
ATCTTCATGGCCTTAACCTCTCTCTTCCGGCAACAACATCACTCCAAACCTAAAACTC
ATTTCTCCCAATCCCCTTCTTCTTCTGGCCATCCGACCGTCACTCTCGACCTCACAACCT
CCTCCTCGTCGAGCAACCGTTCTTATCAATGCTCAATAGATTCAGCTCTCCTCCAAGTA
ATGTCTCAGCATCTAATAGTTATCCTTCAACCAATCTCAACTTTTCAAACAACCAACA
CATTGATGAATTGGGGTGGTGGTGAATCCAGTGATCAATACCGTGCAGCTTACGGCA
ACATTAACACCCATCAGCAATCACCTTACCACAAAATCATTCAAACCCGAACCGCCGGGT
CATCTTTGATCCGTTTGGAGATCATCTTCATCACATTCTCCACAAATAAATCTTGATC
ATATCGGAATCAAGAACATCATCAGTCAACCAAGTGCCATCTTTACCGGCTGAAACAATCA
AGGCAATCACGACAGATCCAAGTTTCCAATCGGCTTTGGCGACAGCTCTATCTTCCATCA
TGGGCGGCGATTAAAGATTGATCACAATGTGACTAGAAATGAAGCTGAGAAGAGCCCTT

AAAGAGAATTGTTATATATATGTTCTTATATACTCAGTACATTGGTAAATGGGTTTAGAC
 TTTCAGTATTTCTAGTTTCATCTATATATTGGTTGTTAATCACAAGTTATTTTGTG
 TTGGAGTTTATGGAACATAATGTGTACATATGAACTTTAGAACGAATAAAATAAACTTGG
 AATTCCTTTTAAAAAAAAAAAAAAAAA

>G2110 Amino Acid Sequence (conserved domain in AA coordinates:239-298)

MEKDDFLRSGHGREESHDEMRKLDSSHDDSHQEHDIIRSKLDSTKVEMDEAKEENRRLK
 SSLSKIKKDFDILQTYNQLMAKHNEPTKFQSKGHHQDKGEDEDREKVNREELVSLSLG
 RRLNSEVPSGSNKEEKNKDVEEAEGDRNYDDNEKSSIQGLSMGIEYKALSNPNEKLEIDH
 NQETMSLEISNNKIRSONSFQKNDGDDHEDEDEILPQNLVKKTRVSVSRRCETPTMND
 GCQWRKYGQKIAKGNPCPRAYYRCTIAASCPVRKQVQRCSEDMSILISTYEGTHNPLPM
 SATAMASATSAASMLLSGASSSSAAADLHGLNFSLSGNNTPKPKTHFLQSPSSSGHP
 TVTLDLTSSSSQOPFLSMLNRFSSPPSNVSRNSYPTNLNFSNNTNLTLMNWGGGNGPS
 DQYRAAYGNINTHQQSPYHKIIQTRTAGSSFDPFGRSSSSHSPQINLDHIGIKNIIHQV
 PSLPAETIKAITDPSPFQALATALSSIMGGDLKIDHNVTRNEAEKSP*

>G2442 (71..997)

TCGACCAATTTAGACCATTCCAAATTCGTCGTCCTTTTCTCTGTGTAGTCTAATTATATA
 TTACAAGTAGATGAATTGGTTACCTGAAGCTGAAGCTGAGGAGCACTTGAAAGGTATTCT
 CTCTGGTGATTTCTTTGATGGTCTCACCATCACCTTGATTGCCCACTTGAGACATCGA
 TTCCACCAATGGTGAGGGAGATTGGGTCGCCAGGTTTCAAGACCTTGAGCCTCCTCCCTT
 GGATATGTTCCCTGCTTTGCCTTCTGACCTCACCTCTTGTCCCAAGGGCGCCGCTCGTGT
 GCGGATTCCTCAACATGATTCTGCTTGAAGCAGTCTGTTCTTCTGAAGCCTTGTC
 CGGCATTAATAGCACTCCCCACCAATCTTCAGCTCCTCTGATATCAAAGTTTCATATCT
 ATTTGAGTCTCTAACTCCAGTGTCTAGTTCTCGAGAACAGTTATGGTTCTCTCTCCACCCA
 AAACTCCGGATCTCAGAGATTGGCTTTCCCTGTGAAAGGCATGAGAAGCAAGCGCAGACG
 CCCCACAACAGTGAGACTTAGCTACCTTTTCCCTTTGAACCCAGAAAGTCAACTCCGGG
 TGAATCAGTAACCGAGGGTTACTATTCTTCTGAGCAACATGCCAAGAAGAAGCGCAAGAT
 TCATCTGATCACCCACACCGAGTCTTCCACTTTGGAGTCAAGTAAGTCGGATGGGATAGT
 CCGGATATGCACTCATTGTGAGACAATCACGACCCACAGTGGAGGCAAGGACCCAGTGG
 ACCCAAGACCCTCTGCAACGCTTGCAGGAGTCCGGTTCAAATCTGGTCGCCTAGTTCCAGA
 ATACCGGCCAGCCTCAAGCCCGACCTTCATCCCATCTGTGCATTCAAACCTCACACAGGAA
 GATCATTGAGATGAGAAAGAAGGACGACGAGTTTGATACCAGCATGATTTCGAGTGATAT
 CCAGAAGGTAAAGCAGGGGAGGAAGAAATGGTATAAAAGTA

>G2442 Amino Acid Sequence (domain in aa coordinates: 220-246)

MNWLPEAEAEHLKGLSGDFDGLTNHLDCPLEDIDSTNGEGDWVARFQDLEPPPLDMF
 PALPSDLTSCPKGAARVRIPNMIPALKQSCSSEALSGINSTPHQSSAPPDIKVSYLQFQ
 LTPVSVLENSYGLSTQNSGSQRLAFPVKGMRSKRRRPTTVRLSYLFPFEPKSTPGESV
 TEGYYSSEQHAKKKRKHILITHESSTLESSKSDGIVRICTHCETITTPQWRQGPSGPKT
 LCNACGVRFKSGRLVPEYRPASSPTFIPSVHSNSHRKIIEMRKKDDEFDTSMIRSDIQKV
 KQGRKKMV*

>G1051 (66..1031)

CCTGTAAATTAGATTGCTTTCTTTGGTAATCTTTGGATCAAGATCCATCTATTTTTT
 CTTCAATGGCACAACTCCCTCCTAAAATCCCCAACATGACACAACATTGGCCTGATTTCT
 CTTCCCAAAAGCTCTCTCCTTTCTCTACCCCAACCGCAACCGCTGTGCGCCACCGCTACAA
 CCACCGTACAAAACCCCTCATGGGTCGACGAATTCCTCGACTTCTCAGCGTCTCGCCGTG
 GCAACCACCGTCGTTCCATCAGCGACTCTATCGCATTCCTCGAAGCTCCAACAGTCAGCA
 TCGAAGACCACCAATTGACAGGTTTCGATGACGAACAGTTTCATGTCGATGTTACCGACG
 ACGACAACCTTCATAGCAATCCTTCCCATATCAACAACAAAATAACAATGTGGGGCCCA
 CGGGATCTTCCCTGAACACATCCACGCGTCCAATAGCTTCAACGACGATAACAAAGAAT
 TACCACCGTCCGATCAATAACATGAACAATAATATCAACAACACTATAACGATGAAGTCC
 AAAGCCAATGCAAGATGGAGCCAGAAGATGGTACGGCGTCAATAACAATTCGGGTGATA
 GCTCCGGCAACCGGATTCTCGATCCCAAAAGGGTTAAGAGAATATTAGCAATCGGCAAT
 CAGCACAGAGATCAAGGGTGAGGAACTGCAATACATATCAGAGCTCGAACGTAGCGTCA
 CTTGCTGTCAGGCGGAAGTGTCTAGTGTATCGCCAAGAGTTCATTCTTGATCATCAAC
 GTTTGCTTCTTAACGTTGACAACAGCGCTCTCAAGCAACGAATCGCTGCTTTATCTCAAG
 ACAAGCTTTTCAAAGACGCACATCAAGAAGCATTGAAGAGAGAAATAGAGAGACTTCGAC
 AAGTGTATAATCAACAAAGCCTCACGAATGTGGAAATGCAATCATTATCGGCGACCG
 GAGCCGGTGCTACTCCGGCCGTCGACATCAAGTCGTCCTGTTGAAACAGAGCAGCTCCTCA

ATGTCTCATAAATTAACCATCATGCATCATCAACATTTCTCTCTTTTAGCTTCTTGG
 CAAAAGTTCTTGACTATAAAATCTCTTTTCGGGTGAAGAAATTCAGGAGATATACATTTTTT
 ATTCTAATCACATTGTTTTTAAGTTGTGATGAATTCAGTTTGATGTATCTTATTTATTTT
 GTTTATGTCGTCTTTTTTTCTTGGGGTTGATGGAAGGAATCATCAATTGTTGTTGTAC
 AAAGAACTAGTTGAATTTTTTTTTTTTTTTT

>G1051 Amino Acid Sequence (domain in AA coordinates 189-250)
 MAQLPPKIPNMTQHWPDPFSSQKLSPFSTPTATAVATATTTVQNPSWVDFLDFSASRRGN
 HRRSISDSIAFLEAPTVSIEDHQFDRFDDEQFMSMFTDDDNLHSNPSHINNKNNVGPTG
 SSSNTSTPSNSFNDDNKELPSPDHNMNNNNNNYNDVQSQCMEPEDGTASNNNSGDS
 GNRILDPKRVKRI LANRQSAQRSVRKLYISELERSVTSLOAEVSVLSPRVAFLDHQRL
 LLNVDNSALKQRI AALSQDKLFKDAHQEALKREIERLRQVYNQQLTNVENANHL SATGA
 GATPAVDIKSSVETEQLLNVS*

>G1052 (138..1127)

TGATCATCTAAACTTTCAATTTCTCTCTTGATCCTCACTTGAATTTTTTGTGTTTCTC
 TCAAACTTTTGATCCTTTCTTTGTTTTTCATTGACCTCTTACAAAAAATCTGGTGTG
 CCATTAAATCTTTATTAATGGCACAACCTCTCCGAAAATCCCAACCATGACGACGCCAA
 ATTGGCCTGACTTCTCTCCAGAACTCCCTTCCATAGCCGCAACGGCGGCAGCCGCAG
 CAACCGCTGGACCTCAACAACAAAACCTTCATGGATGGATGAGTTTCTCGACTTCTCAG
 CGACTCGCGTGGGACTACCGTCTGTCTATAAGCGACTCCATTGCTTTCCTGAACCAC
 CTCTCTCCGGCGTCGGAACCACTTCGATAGGTTTGACGACGAGCAATTCATGTCCA
 TGTTCAACGACGACGTACACAACAATAACCAATCATCATCATCACAGCATCAACG
 GCAATGTGGGTCCACGCGTTCATCTCCAACACCTCCACGCGTCCGATCATAATAGCC
 TTAGCGACGACGACAACAACAAGAAGCACCACCGTCCGATCATGATCATCATGAGACA
 ATAATGTAGCCAATCAAAACAACGCGCGCGTAACAATTACAACGAATCAGACGAGGTCC
 AAAGCCAGTGCAAGACGAGCCACAAGATGGTCCGTCGGCGAATCAAACTCCGGTGGAA
 GCTCCGGTAATCGTATTCACGACCTAAAAGGGTAAAAGAAATTTAGCAAATAGGCAAT
 CAGCAGAGATCAAGGGTGAGGAAATGCAATACATATCAGAGCTTGAAAGGAGCGTTA
 CTTCAATGCGAGACTGAAGTGTAGTGTATCGCCAAGAGTTGCGTTTTTGGATCATCAGC
 GATTGCTTCTCAACGTGCGACAATAGTGTATCAAGCAACGAATCGCAGCTTTAGCACAAG
 ATAAGATTTTCAAAGACGCTCATCAAGAAGCATTGAAGAGAGAAATAGAGAGACTTCGAC
 AAGTATATCATCAACAAAGCCTCAAGAAGATGGAGAATAATGTCTCCGATCAATCTCCGG
 CCGATATCAAACCGTCCGTTGAGAAGGAACAGCTCCTCAATGTCTAAAGCTGTTCGTTCA
 CTAAGATCTTTCTTTTCATGGCGAAAAGATTCTTGACTATAAAACCTCTTTGTGTCAAGA
 AATTAATTTATCAAAGAAGATGGCCTTTTTTATTTGATCTAATCACATTTTTTTAAGTTG
 TGATGAATTTGCTTTTGATGTATCTGTTTTTTTTTTTTTTTTT

>G1052 Amino Acid Sequence (domain in AA coordinates 201-261)
 MAQLPPKIPTMTTPNWPDPFSSQKLPSIAATAAAAATAGPQQNPSWMDFLDFSATRRGT
 HRRSISDSIAFLEPPSSGVGNHFDREFDDEQFMSMFDNDVHNNNNHHHHHSINGNVGPT
 RSSNTSTPSDHSLSDDNNKEAPPSPDHHDHMDNNVANQNNAAAGNNYNESDEVQSCKT
 EPQDGPSANQNSGGSSGNRIHDPKRVKRI LANRQSAQRSVRKLYISELERSVTSLOTE
 VSVLSPRVAFLDHQRL LLNVDNSAIKQRI AALAQDKIPKDAHQEALKREIERLRQVYHQQ
 SLKKMENNVSQSPADIKPSVEKEQLLNVS*

>G1079 (1..1995)

ATGGGTTGTGCTGCTTCAAGAATTGATAATGAAGAAAAGGTTTTAGTGTGTAGGCAGAGA
 AAGAGGCTAATGA AAAAGTTATTAGGGTTTCAGGGGAGAATTTGCAGATGCACAGTTGGCT
 TATCTTAGAGCTTTGAGGAACACTGGTGTTACTCTTAGGCAATTCAGTGAGTCTGAGACC
 TTGGAGCTTGAAAACACTAGTTATGGTTTAAGTTTGCCTTTGCCTCCTTCGCTCCTCCT
 ACATTGCCTCCTTCACCTCCACCACCTCCTCCATTTAGCCCGGATTTGAGAAATCCTGAG
 ACTAGTCATGACTTGGCTGATGAGGAGGAAGAGGGTGAAAATGATGGTGGTAATGATGGA
 AGTGGTGCAGCTCCTECGCCTCCATTGCCGAATCTTGGAACATTTGGAACCTTTTGAG
 TCATTGAGCTGCATAGTCATCCAAATGGTGACAATGTAGTTACACAAGTTGAACTGAAG
 AAGAAACAACAAATTCAGCAAGCTGAAGAGGAAGATTGGGCGGAGACGAAGTCTCAATTT
 GAGGAAGAAGATGAGCAACAAGAAGCAGGAGGTACTTGCCTTGATTTGAGTGTTTCATCAA
 ATAGAGGCTGTGTAGTGGCTGTAAACATGAAGAAGCCAGTCGTCTGAAGTTTAAGCTGGGA
 GAAGTTATGGACGGTAACCTCATCTATGACAAGCTGCTCCGGTAAAGATCTTGAGAAAAC
 CATGTGACTGATTGTAGAATCAGGAGGACCTTAGAAGGAATCATCAGAGAGTTGGATGAT
 TATTTCTTAAAGCATCGGGTTGCGAGAAGGAGATAGCTGTGATAGTAGACATCAACAGT

AGGGATACTGTTGATCCTTTTCAGGTACCAGGAAACAAGAAGGAAGAGAAGCAGCTCGGCA
 AAGGTATTTCAGTGCATTGTTCATGGAGTTGGTCTTCAAAGTCTCTTCAGTTGGGCAAAGAT
 GCTACAACAAGCGGGACTGTTGAACCTGTAGGCTGGAGCTCACTGCAGCACACTTGAG
 AAGCTATACACAGCTGAGAAGAACTTTACCAGCTAGTCAGAAACAAGAGATTGCCAAA
 GTGGAGCATGAGAGGAAGTCTGCATTACTGCAAAAGCAAGATGGGGAAACCTATGATTTG
 AGCAAAATGGAGAAAGCACGCTTGTCTTTGGAGAGTTTGAAACCGAGATACAGCGTCTA
 GAAGATTCCATAACTACAACACGCTCATGTTTGTCTTAACCTTGATCAATGATGAGCTGTAT
 CCGCAGCTAGTTGCTTTAACTTCAGGGCTAGCACAGATGTGGAAAACAATGCTCAAGTGT
 CATCAAGTTCAAATTCATATATCCAGCAACTGAACCATCTTCCGATTACCCGAGTATA
 GATCTCAGTTTCGGAATACAAACGCCAGGCGGTAAATGAAGTAGAGACCGAGGTTACTTGC
 TGGTACAATAGCTTTTGCAGTTAGTAAATTCACGCGAGAATACGTGAAAACACTCTGT
 ACGTGGATCCAACCTACTGATCGCCTCTCTAACGAAGACAACCAAAGAAGTAGCTTGCCT
 GTTGTCTGCTCGTAAGCTCTGCAAAGAGTGGCAGCTTGAATACAACCTGCGTAGGAAATGC
 AATAAACTTGAGAGGAGGCTTGAGAAAGAGCTAATTTCACTGGCTGAGATTGAAAGAAGG
 CTCGAGGGGATTTTAGCAATGGAAGAGGAGGAAGTAAGCTCAACGAGTTTGGGCTCTAAG
 CATCCGTTGTCAATCAAACAAGCCAAGATCGAAGCCTTGAGAAAACGAGTGGATATTGAG
 AAACTAAGTACTTAACTCGGTGAGGTTAGTAAGAGAATGACACTAGACAACCTCAAA
 TCAAGCCTTCCAATGTCTTTAGATGTTGACTGCTCTAGCTAATGTCTTTGCCAATGGG
 TTTGAATCCGTTAATGGCCAAACCGGTACAGATGTTTCCGACACATCCCAACATTCCGAT
 GAATCTCAACCCTAA

>G1079 Amino Acid Sequence (conserved domain in AA coordinates:1-50)

MGCAASRIDNEEKVLVCRQRKRLMKLLGFRGEFADAQLAYLRALRNTGVTLRQFTBSET
 LELENTSYGLSLPLPPSPPTLPSPPPPPFSPDLRNPETSHDLADEEEEEENDGGNDG
 SGAAPPPPLPNSWNIWNPFESELELHSHPNGDNVVTQVELKKKQQIQQAEEEDWAETKSQF
 EEDEEQQEAGGTCLDLVSHQIEAVSGCNMCKPRRLKFKLGEVMDGNSSMTSCSGKDLEKT
 HVTDCRIRRTLEGIIRELDDYFLKASGCEKEIAVIVDINSRDTVPFRYQETRRKRSSSA
 KVFSALSWSWSKSLQLGKDATTSGTVEPCRPGAHCSTLEKLYTAEKKLYQLVRNKEIAK
 VEHERKSALLQKQDGETYDLSKMEKARLSLESLETEIQRLSDSITTTSRCLLNLINDELY
 PQLVALTSGLAQMWMKMLKCHQVQIHISSQLNHLDPDYSIDLSSEYKRQAVNELETEVTC
 WYNSFCKLVNSQREYVKTLCTWIQLTDRLSNEDNQRSSLPVAARKLCKEWQLEYNLRRKC
 NKLERLEKELISLAEIERRELEGILAMEEEEEVSSTSLGSKHPLSIKQAKIEALRKRVDIE
 KTKYLNSEVSKRMTLDNLKSSLPNVFQMLTALANVFANGFESVNGQTGTDVSDTSQHSD
 ESQP*

>G1335 (56..667)

TTTTTTTTTAAAGATTTAGAGAGAAAAGTGAGTTATTAAGAGATTCCAATCAAAATGAG
 CGGAGACAACGGCGGTGGTGAGAGCGCAAAGGCTCCGTCAAGTGGTTTGATACCCAGAA
 GGGTTTCGGCTTCATCACTCCTGACGACGGTGGCGACGATCTCTTCGTTCCAGTCCCTC
 CATCAGATCTGAGGGTTTCCGTAGCCTCGCTGCCGAAGAAGCCGTAGAGTTCGAGGTTGA
 CATCGACAACAACAACCGTCCCAAGGCCATCGATGTTTCTGGACCCGACGGCGCTCCCGT
 CCAAGGAAACAGCGGTGGTGGTTTCATCTGGCGGACGCGCGGTTTCGGTGGAGGAAGAGG
 AGGTGGACGCGGATCTGGAGGTGGATACGGCGGTGGCGGTGGTGGATACGGAGGAAGAGG
 AGGTGGTGGTGGAGGAGGACGACTGCTACAAGTGTGGTGGAGCCGGTCACATGGCGAG
 AGACTGTTCTGAAGGCGGTGGAGGTTACGGAGGAGGCGCGGTGGCTACGGAGGTGGAGG
 CGGATACGGCGGAGGAGGTGGTGGTTACGGAGGTGGTGGCCGTGGAGGTGGTGGCGGCGG
 GGGAGCTGCTACAGCTGTGGCGAGTCGGGACATTTCCGACGGGATTGCACACGCGGTGG
 ACGTTAAAACCAACGCCGTTACGCGGTGGAGAAGAGTGAGTTGGTTATCTCACAAGTGA
 TCGGTTCTTTCTCCCGCCGCTTCTATCTCTATTATCCACTTTTGTCTATTATGATG
 GATCTCTATCTTTGTAGTTGGTTTTTCTTGATGGTTTCGATTAGGACTCTTCTTTTG
 GTTTTGCTACTTATGGTTGGTTTTTATTATGGTACTTGTGATATGGGTGAAATGCTCTAC
 TTGTTGCTCTGTTTCAAGTGTTTATAATATGCGAACAATATTCTGGGTTTTGTTTCAAA
 AAAAA

>G1335 Amino Acid Sequence (domain in AA coordinates: 24-43, 131-144, 185-203)

MSGDNGGGERRKGSVKWFDTPQKGFITPDDGGDDLFDVHQSSIRSEGFRLAAEEAVEFE
 VEIDNNNRPKAIDVSGPDGAPVQGNSSGGSSGGRRGGFGGGRRGGSGGGYGGGGGGYGG
 RGGGGRRGSDCYKCEPFGHMARDCEGGGGYGGGGGGYGGGGYGGGGGGYGGGGRRGGGG
 GGGSCYSCGESGHFARDCTSGGR*

>G157 (31..621)

GGGCATAACCCCTTATCGGAGATTTGAAGCCATGGGAAGAAGAAAAATCGAGATCAAGCGA
ATCGAGAAACAAAGCAGTCGACAAGTCACCTTCTCAAACGACGCAATGGTCTCATCGAC
AAAGCTCGACAACCTTTCGATTCTCTGTGAATCCTCCGTCGCTGTTGTGTCGTATCTGCC
TCCGGAAAACTCTATGACTCTTCTCCGGTGACGACATTTCCAAGATCATTGATCGTTAT
GAAATACAACATGCTGATGAACCTTAGAGCCTTAGATCTTGAAGAAAAAATTCAGAATTAT
CTTCCACACAAGGAGTTACTAGAAACAGTCCAAAGCAAGCTTGAAGAACCAAATGTCGAT
AATGTAAGTGTAGATTCTCTAATTTCTCTGGAGGAACAACTTGAGACTGCTCTGTCCGTA
AGTAGAGCTAGGAAGGCAGAACTGATGATGGAGTATATCGAGTCCCTTAAAGAAAAGGAG
AAATTGCTGAGAGAAGAGAACAGGTTCTGGCTAGCCAGATGGGAAAGATAACGTTGCTG
GCAACAGATGATGAGAGAGGAATGTTTCCGGGAAGTAGCTCCGGCAACAAAATACCGGAG
ACTCTCCCGCTGCTCAATTAGCCACCATCATCAACGGCTGAGTTTTACCTTAAACTCAA
AGCCTGATTATATAAAGATAAATTTGTATATTATAAAAAGCTGTGTAATCTCAAA
CCTTTTATCTTCTCTAGTGTGGAATTTAAGGTCAAAAAGAAAACGAGAAAGTATGGATC
AGTGTGTACCTCCTTCGGAGACAAGATCAGAGTTTGTGTGTTTGTGTCTGAATGTACGG
ATTGGATTTTTAAAGTTGTGCTTTCTTTCTTCAAAAAAAAAA

>G157 Amino Acid Sequence (domain in AA coordinates: 2-57)

MGRRKIBIKRIENKSSRQVTFSKRRNGLIDKARQLSILCESSVAVVVVSASGKLYDSSSG
DDISKIIDRYEIQHADELRLALDLEEKIQNYLPHKELLETVQSKLEEPNVNDVSVDSLISL
EEQLETALSVSRRKAELMMEYIESLKEKEKLLREENQVLASQMGKNTLLATDDERGMFP
GSSSGNKIPETLPLLN*

>G1895 (1..954)

ATGAATAACCAATCTGTTACTGACAATACAAGTCTTAAGCTGTCATCTAATCTTAACAAC
GAGTCAAAAGAAACATCTGAGAACAGTGATGACCAACACAGCGAGATCACAACAATTACA
TCGGAAGAAGAGAAAACAACCTGAAGTGAAGAAACAGACAAGATTCTTCCATGTCCGAGA
TGCAACAGCGCAGACACCAAATCTGTTACTACAACAACCTACAACGTTAACCAGCCACGT
CACTTCTGTAGAAAATGCCAGAGGTATTGGACCGCTGGTGGATCCATGAGGATCGTCCCG
GTTGGCTCAGGCCGTCGCAAGAACAAGGGATGGGTTTCTTCAGACCAGTACCTGCACATC
ACTTCCGAGGATACTGACAATTACAATAGCTCCTCAACAAAGATTCTAAGCTTCGAGTCT
TCGGACTCTTTGGTAACTGAGAGGCCCTAAGCATCAATCAAACGAAGTGAAGATAAACGCT
GAACCTGTTTCAACAAGAACCAACAACCTTCAAGGGTTACTTCTCCCCAAGCATCCCCT
GTTTCGGCTCCTTGGCCTTACCAATACCTTCCAAACCTAGTTTCTACCACATGCCCGTC
TACTGGGGCTGCGCGATACCGTTTGGTCTACCCTCGACACTTCTACATGTCTTGGGAAA
AGGACAAGAGACGAAACTTCTCATGAACTGTTAAAGAGAGTAAAAATGCTTTTGAGAGA
ACAAGCTTGCTTTTGAATCTCAGAGCATCAAAAATGAAACAAGTATGGCTACAAATAAC
CATGTGTGGTATCCAGTACCGATGACCCGCGAGAAGACACAAGAATTCAGCTTTTTCAGT
AATGGAGCTGAAACAAAGAGCAGCAACAACAGATTCTGTCCTGAAACGTATCTTAACCTG
CAAGCAAACCTGCAGCCATGGCAAGATCTATGAACCTTCAGAGAGAGCATATAA

>G1895 Amino Acid Sequence (domain in AA coordinates: 55-110)

MNNQSVTDNTSLKLSSNLNNEKSEKSENSDDQHSEITITSEEKTELKPKDKILPCPR
CNSADTKFCYNNYNNVNPQRHFCRKCQRYWTAGGSMRIVPVGSGRRKNKGWSSDQYLHI
TSEDTDNYSSTKILSFESSDSLVTTERPKHQSNVVKINAEPVVSQEPNNFQGLLPQASP
VSPWPYPYPPNPSPFYHMPVYWGCAIPVWSTLDTSTCLGKRTRDETSHETVKESKNAFER
TSLLESQSIKNETSMATNNHVWYPVPMTREKTQEFSSFSNGAETKSSNNRFPVETYLNL
QANPAAMARSMNFRESI*

>G1900 (1..897)

ATGCTGGAAACTAAAGATCCTGCGATAAAGCTCTTTGGTATGAAAATTCCTTTCCCGACG
GTTTTAGAGGTTGCTGATGAAGAAGAAGAAAAGAACCAAAACAAGACATTAAGTATCAA
TCGGAGAAAGACAAACCCCTAAAGAAACCAACCAAGATTCTTCCATGTCCAAGATGCAAC
AGCATGGAGACTAAGTTCTGTACTACAACAACCTACAACGTAACCAACCTCGCCATTTT
TGTAAGCTTGTGAGAGATATTGGACCTCAGGTGGGACCATGAGAAGTGTCCAATCGGA
GCAGGACGGCGCAAGAACAAGAACAACCTACCAACTTCACATTACCACCATGTGACTATC
TCCGAAACAAATGGTCCGGTCTTAGTTTCAGCCTCGGAGATGATCAAAGGTCTCGAGT
AATAGTTTGGTAAATCAAAGCTAGTTGCTAGGATAGAGAACAATGACGAGCGCTCTAAT
AACAACACTTCGAACGGTTTGAATGTGTTTCCGGGAGTTTCGTGGCCGTACACGTGGAAT
CCTGCGTTTTACCCGGTTTACCCTTATTGGAGCATGCCAGTGTTGTCTTCTCCGGTAAGT
TCAAGTCCTACTTCTACTCTTGGTAAGCATTGAGAGACGAAGACGAGACGGTGAAGCAA
AAACAGAGGAATGGATCTGTATTGGTTCCAAAGACTTTGAGAATTGATGATCCTAATGAA

GCTGCAAAGAGTTCGATATGGACAACACTTGGGATCAAGAACGAAGTTATGTTCAATGGG
 TTTGGTTCGAAGAAAGAGGTTAAGCTCAGTAACAAAGAAGAAACAGAGACCTCACTTGTT
 CTTTGTGCAAACCTGCTGCGTTATCAAGATCAATCAATTCCATGAGCAGATGTGA
 >G1900 Amino Acid Sequence (domain in AA coordinates: 54-106)
 MLETKDPAIKLFGMKIPFPTVLEVADEEEKNQNKLTLDQSEKDKTLKKPTKILPCPRCN
 SMETKFCYYNNYNVNQPRHFCKACQRYWTSGGTMRSPVIGAGRRKNKNSPTSHTYHHVTI
 SETNGPVLSPSLGDDQKVSSNRFNQKLVARIEENNDERSNNNTSGLNCFPGVSWPYTWN
 PAFYPVYPYWSMPVLSSPVSSSPTSTLGKHSRDEDETVKQKQRNGSVLVPKTLRIDDPNE
 AAKSSIWTTTLGIKNEVMFNGFGSKKEVKLSNKEETETSLVLCANPAALSRSINFHEQM*
 >G2007 (1..861)
 ATGGGAAGGCAGCCATGTTGTGACAAGCTCATGGTGAAGAAGGGCCGTGGACGGCGGAG
 GAAGACAAGAACTGATAAACTTTATCTTGACCAACGGCCACTGTTGCTGGAGGGCTTGT
 CCGAAGCTGGCCGCTCTCCGTCGCTGTGGGAAGAGCTGCCGTCTACGGTGGACCAATTAT
 CTCCGACCTGACTTGAAGAGAGGTCTTCTCTCCGACGCCGAGGAACAGCTTGTATCGAC
 CTTTATGCTCTTCTCGGCAACAGATGGTCCAAGATCGCTGCAAGATTACCAGGAAGAACA
 GACAACGAAATAAAAAATCATTTGGAATACTCATATCAAGAAGAAGCTCCTTAAGATGGAA
 ATCGATCCTTTCGACCCATCAACCTTTAAACAAAGTATTTACCGATACAAACTTAGTCGAT
 AAATCTGAACTTTCATCGAAAGCCGACAATGTAAATGATAATAAAATCGTAGAGATCGAT
 GGGACAACGACAAATACAATAGATGATAGCATTATCACTCATCAAAATAGTTCAAATGAT
 GATTATGAATTACTTGGTGATATAATTCTATAATTATGGAGATTTATTTAATATTCTATGG
 ACCAACGATGAACCTCTCTAGTCGATGATGCATCATGGAGCAATCATAACGTTGGTATT
 GGAGGAACAGCTGCAGTTGCAGCCTCAGACAAGAACAACACTGCTGCCGAGGAAGATTTTC
 CCGGAAAGATCATTTGAAAAACAGAACGGCGAAAGTTGGATGTTCTTGGATTATTGCCAA
 GAATTTGGTGTTGAAGATTTTGGGTTTCGAGTGTTACCATGGTTTTGGTCAAAGCTCCATG
 AAGACGGGTCACAAGGACTAG
 >G2007 Amino Acid Sequence (domain in AA coordinates: TBD)
 MGRQPCCDKLMVKGPWTAEDDKKLINFILTNHCCWRALPKLAGLRRCGKSCRLRWNTY
 LRPDLKRGLLSDAEEQLVIDLHALLGNRWSKIAARLPGRDNEIKNHWNTHIKKLLKME
 IDPSTHQPLNKVFTDTNLVDKSETSSKADNVNDNKIVEIDGTTTTNTIDDSIITHQNSSND
 DYELLDGI IHNHYGDLFNILWTNDEPPLVDDASWSNHNVGIGGTAAVAASDKNNTAAEEDF
 PERSFEKQNGESWMFLDYCQEFGEVDFGFECHYHGFQSSMKTGHKD*
 >G214 (238..2064)
 TGAGATTTCTCCATTCCGTAGCTTCTGGTCTCTTTTCTTTGTTTCATTGATCAAAAGCA
 AATCACTTCTTCTTCTTCTTCTTCTCGATTCTTACTGTTTTCTTATCCAACGAAATCTG
 GAATTAAAAATGGAATCTTTATCGAATCCAAGCTGATTTTGTTCCTTTTATTGAATCATC
 TCTCTAAAGTGGAATTTTGTAAAGAGAAGATCTGAAGTTGTGTAGAGGAGCTTAGTGATG
 GAGACAAATTTCGTCTGGAGAAGATCTGGTTATTAAGACTCGGAAGCCATATACGATAACA
 AAGCAACGTGAAAGGTGGACTGAGGAAGAACAATAAGATTGATTGAAGCTTTGAGGCTT
 TATGGTAGAGCATGGCAGAAAGATTGAAGAACATGTAGCAACAAAACCTGCTGTCCAGATA
 AGAAGTCACGCTCAGAAATTTTCTCCAAGGTAGAGAAAGAGGCTGAAGCTAAAGGTGTA
 GCTATGGGTCAAGCGCTAGACATAGCTATTCTCTCTCCACGGCCTAAGCGTAAACCAAAC
 AATCCTTATCTCGAAAGACGGGAAGTGGAACGATCCTTATGTCAAAAACGGGTGTGAAT
 GATGGAAAAGAGTCCCTTGGATCAGAAAAAGTGTGCGATCCTGAGATGGCCAATGAAGAT
 CGACAACAATCAAAGCCTGAAGAGAAAACTCTGCAGGAAGACAACCTGTTTCAAGTTGTTTC
 ACTCATCAGTATCTCTCTGCTGCATCCTCCATGAATAAAAGTTGTATAGAGACATCAAAC
 GCAAGCACTTTCCGCGAGTTCTTGCCTTCACGGGAAGAGGGAAGTCAGAATAACAGGGTA
 AGAAAGGAGTCAAACCTCAGATTGAATGCAAAATCTCTGGAAAACGGTAATGAGCAAGGA
 CCTCAGACTTATCCGATGCATATCCCTGTGCTAGTGCCATTGGGGAGCTCAATAACAAGT
 TCTCTATCACATCCTCCTTCAGAGCCAGATAGTCATCCCCACACAGTTGCAGGAGATTAT
 CAGTCGTTTCTTAATCATATAATGTCAACCCTTTTACAAACACCGGCTCTTTATACTGCC
 GCAACTTTTCGCTCATCATTTTGGCCTCCCGATTCTAGTGGTGGCTCACCTGTTCCAGGG
 AACTCACCTCCGATCTGGCTGCCATGGCCGAGCCACTGTTGAGCTGCTAGTGCTTGG
 TGGGCTGCCAATGGAATTATTACCTTTATGTGCTCCTCTTAGTTTCAAGGTGGTTTCACTAGT
 CATCCTCCATCTACTTTTGGACCATCATGTGATGTAGAGTACACAAAAGCAAGCACTTTA
 CAACATGGTTCTGTGCAGAGCCGAGAGCAAGAACAACCTCCGAGGCATCAAAGGCTCGATCT
 TCACTGGACTCAGAGGATGTTGAAAATAAGAGTAAACCAGTTTGTGATGAGCAGCCTTCT
 GCAACACCTGAGAGTGATGCAAAGGGTTCAGATGGAGCAGGAGACAGAAAACAAGTTGAC

CGGTCCTCGTGTGGCTCAAACACTCCGTCGAGTAGTGATGATGTTGAGGCGGATGCATCA
 GAAAGGCAAGAGGATGGCACCAATGGTGAGGTGAAAGAAACGAATGAAGACACTAATAAA
 CCTCAAACCTCAGAGTCCAATGCACGCCGAGTAGAATCAGCTCCAATATAACCGATCCA
 TGGAACTCTGTGTCTGACGAGGGTGAATTGCCTTCCAAGCTCTCTTCTCCAGAGAGGTA
 TTGCCGCAAGAGTTTACATATCGAGAAGAACACAGAGAGGAAGAACAACAACAAGAA
 CAAAGATATCCAATGGCACTTGATCTTAACCTCACAGCTCAGTTAACACCAGTTGATGAT
 CAAGAGGAGAAGAGAAACACAGGATTTCTTGGAATCGGATTAGATGCTTCAAAGCTAATG
 AGTAGAGGAAGAACAGGTTTTAAACCATACAAAAGATGTTCCATGGAAGCCAAAGAAAGT
 AGAATCCTCAACAACAATCCTATCATGTGGAACAGAAAGATCCCAAACGGATGCGG
 TTGGAACTCAAGCTTCCACATGAGACTCTATTTTCATCTGATCTGTTGTTGTAATCTG
 TTTTAAAGTTTTCAAGACCCTGCTACATTTCTTTTCTTTTGAAGGCTTTGTATTTGT
 TTCCTTGTCATAGTCTTCTGTAAACATTTGACTCTGTATTATTCAACAAATCATAAACT
 GTTAACTTTTTTTTTTCCA

>G214 Amino Acid Sequence (domain in AA coordinates: 22-71)
 METNSSGEDLVIKTRKPYTITKQREERWTEEEHNRFIELRLYGRAWQKIEBHVATKTAVQ
 IRSHAQKFFSKVEKEAEAKGVAMGQALDIAIPPPRPKRKPNNPYPRKTGSGTILMSKTGV
 NDGKESLGSEKVSHPMANEDRQQSKPEEKLQEDNCSDFTHQYLSAASSMNKSCIETS
 NASTFREFLPSREEGSQNNRVRKESNSDLNAKSLENGNEQGPQTYPMHIPVLVPLGSSIT
 SSLSHPPSEPDSHPHTVAGDYQSFPHIMSTLLQTPALYTAATFASSFWPPDSSGGSPVP
 GNSPPNLAAMAAATVAASAWWAANGLLPLCAPLSSGGFTSHPPSTFGPSCDVEYTKAST
 LQHGVSQSREQEHSEASKARSSLDSEDEVENKSKPVCHEQPSATPESDAKGSAGDRKQV
 DRSSCGSNTPSSDDVEADASERQEDGTNGEVKETNEDTNKPQTSSESNARRSRISNITD
 PWKSVSDEGRIFAFQALFSREVLQSFYREEHREEEQQQEQRYPMALDLNFTAQLTPVD
 DQEEKRNTGFLGIGLDASKLMSRGRGTGFKPKYKRCMEAKESRILNNNPIIHVEQKDPKRM
 RLETQAST*

>G2155 (63..740)

CTCATATATACCAACCAAACCTCTCTGTCATCTTTATTAAACACAAAATTCAAAAGATT
 AAATGTTGTGCAAGCTCCCTACACAGCGACACTTGACCTCTCTCCCTCCTCTCCCTCCA
 TGGAAACCGTCGGGCGTCCACGTGGCAGACCTCGAGGTTCCAAAACAAACCTAAAGCTC
 CAATCTTTGTGACCATTTGACCTCCTATGAGTCCCTACATCCTCGAAGTGCCATCCGGAA
 ACGATGTGCTGTTGAAGCCGTAAACCGTTTCTGCCGCGGTAAAGCCATCGGCTTTTGCCTCC
 TCAGTGGCTCAGGCTCCGTTGCTGATGTCACTTTGCGTCAGCCTTCTCCGCGAGCTCCTG
 GCTCAACCATTACTTTCCACGGAAAGTTGATCTTCTCTGTCTCCGCCACTTTCTCTCC
 CTCTCTACCTCCTACCTCCTTGTCCCCTCCCGTCTCCAATTTCTTCACCGTCTCTCTCG
 CCGACCTCAGGGGAAAGTCATCGGTGGATTCTGTCGCTGGTCTCTCGTTGCCGCGGAA
 CTGTTTACTTCGTCGCCACTAGTTTCAAGAACCCTTCTATACCGGTTACCTGCTACGG
 AGGAAGAGCAAAGAACTCGGCGGAAGGGGAAGAGAGGGACAATCGCCGCCGGTCTCTG
 GAGGTGGTGGAGAGTCGATGTACGTGGGTGGCTCTGATGTCAATTTGGGATCCCAACGCCA
 AAGCTCCATCGCCGTACTGACCACAAATCCATCTCGTTCAAACAGGGTTTCTTCTTCTT
 TAGATCATCAAGAATCAACAAAAGATTGCATTTTAGATTCTTTGTAATATCATAATTG
 ACTCACTCTTTAATCTCTCTATCACTTCTTCTTTAGCTTTTCTGCAAGTGTCAAACCTCA
 CATATTTGTAGTTTGTATTTGACTATCCCAAGTTTGTATTTTATCATACAAATTTTGC
 CTGTCTCTAATGGTTGTTTTTCTGTTGTATAATCTTATGCATTGTTTATTGGAGCTCCA
 GAGATTGAATGTATAATATAATGGTTTAAAT

>G2155 Amino Acid Sequence (domain in AA coordinates: 18-38)
 MLSKLPTQRHLHLSPSSPSMETVGRPRGRPRGSKNPKAPIFVTIDPPMSPYILEVPSGN
 DVVEALNRFRCRGAIGFCVLSGSGSVADVTLRQPSAPAGSTITFHGKFDLLSVSATFLP
 PLPPTSLSPPVSNFPTVSLAGPQGVIGGFVAGPLVAAGTVYFVATSFKNPSYHRLPATE
 EEQRNSAEGE EEGQSPPVSGGGESMYVGGSDVIWDPNAKAPSPY*

>G234 (106..1035)

CACAACATCATACCCACCAACATATATAATCTTGATCATAGAGAGATAAACAGAGGCCGC
 TATCAAGAACAAGACTAAGAACAAGACTTCACTAGGAGTACAAGTATGGGAAGAGCACCG
 TGTGTGTGACAAAGCAAACGTGAAGAAAGGGCCTTGGTCTCCTGAGGAAGATGCAAACTC
 AAATCTTACATTGAAAATAGTGGCACCGGAGGCAATTGGATCGCTTTGCCTCAAAGATT
 GGTTTAAAGAGATGTGGAAAGAGTTGCAGGCTGAGGTGGCTTAACTATCTTAGACCAAAC
 ATCAAACATGGTGGCTTCTCTGAGGAAGAAGAAAACATCATTTGTAGCCCTTTACCTTACA
 ATTGGTAGCAGGTGGTCTATAATCGCTGCTCAATTGCCGGGACGAACAGACAACGATATA

AAAACTATTGGAACACGAGGCTCAAGAAGAACTCATTAAACAAACGCAAGGAGCTT
CAAGAAGCTTGTATGGAGCAGCAAGAGATGATGGTGATGATGAAGAGACAACACCAACAA
CAACAAATCCAACTTCTTTTATGATGAGACAAGACCAACAATGTTTACATGGCCACTA
CATCATCATAATGTTCAAGTTCAGCTCTTTTCAGAATCAAACCAACTCGTTTTGCGACC
AAGAAGATGTTAAGCCAGTGCTCATCAAGAACATGGTCAAGATCGAAGATCAAGAAGTGG
AGAAAACAAACCTCATCATCATCAAGATTCAATGACAACGCTTTTGATCATCTCTCTTTTC
TCTCAACTCTTGTTAGATCCTAATCATAACCACTTAGGATCAGGAGAGGGTTTCTCCATG
AACTCTATCTTGAGCGCCAACACAACTCTCCATTGCTTAACACAAGTAATGATAATCAG
TGGTTCCGGAATTTCCAGGCCGAAACCGTAAACTTGTCTCAGGAGCCTCCACAAGTACT
TCGGCAGATCAAAGCACTATAAGTTGGGAAGACATAAGCTCTCTTGTATTCTGATTCA
AAGCAATTTTTTTAATTATAATAATATATTCTTAAGATGAAACGTACATCATTATTA
TTAATTGGGGGTACGTAACGTATATATGGAATAACGATCTAGTTTGTTTAAATTTAAAA
>G234 Amino Acid Sequence (domain in AA coordinates: 14-115)
MGRAPCCDKANVKKGPWSPEEDAKLSYIENSGTGGNWIALPQKIGLKRCKGKSCRLRWLN
YLRPNIKHGGFSEEEENIICSLYLITIGSRWSIIAQLPGRTDNDIKNYWNTRLKKLINK
QRKELQEACMEQQEMVMKROHQQQQIQTSFMMRQDQTMFTWPLHHHNQVQVPALFRIKP
TRFATKMLSQCSSRTWSRSKIKNWRKQTSSSSRFNDNAFDHLSFSQLLDPNHNHLGSG
EGFSMNSILSANTNSPLNLTNSNDNQWFGNFQAETVNLFSGASTSTSDQSTISWEDISSL
VYSDSKQFF*

>G361 (54..647)

TCTGTCTCTCTCTCTCTCTTTGTAAATATACATATATAGATAAGCTCACATATATGGCGA
CTGAAACATCTTCTTTGAAGCTCTTCGGTATAAACCTACTTGAAACGACGTCGGTTCAAA
ACCAGTCATCGGAACCAAGACCCGGATCCGGATCAGGATCCGAGTCACGTAAGTACGAGT
GTCAATACTGTTGTAGAGAGTTTGCTAACTCTCAAGCTCTTGGTGGTCACCAAAACGCTC
ACAAGAAAGAGCGTCAGCTTCTTAAACGTGCACAGATGTTAGCTACTCGTGGTTTGCCAC
GTCATCATAATTTTACCCCTCATACCAATCCGCTTCTCTCCGCTTCGCGCCGCTGCCTC
ACCTCTCTCTCAGCCGCATCCTCCGCCGCATATGATGCTCTCTCTCTCTCTCTCGAGTT
CTAAGTGGCTTTACGGTGAACACATGTCGTCACAAAACGCCGTTGGGTACTTTCATGGTG
GAAGGGGAGCTTTACGGAGGTGGCATGGAGTCTATGGCCGGAGAAGTAAAGACTCATGGTG
GTTCTTTGCCGAGATGAGGAGGTTCCGCCGAGATAGTGATCGGAGTAGCGGAATTAAGT
TAGAGAATGGTATTGGGCTGGACCTCCATTTAAGCCTTGGGCCATGAATGATTATAATTT
TGGCCAGTAAAGATCTGTAAATACTACTAGGATTTTATTTTATAGAGTATGTTTTTT
TCCTTAATTTTCGGTTGAAATTGGTGAATATTTTATCTCTTACTTACCAATCTCATATT
TCTATGTATGCGTTTGCTTCACTTTTTTTTTTATATAATTCTTCTGTAAAAAATGCA
ATGTGAGTTTTCTTCCCTATCATCTGTCAAGCTTTGGTTCAATTATTTAGTAATCGAAT
AATATAGGAATAGTGTGAAAG

>G361 Amino Acid Sequence (domain in AA coordinates: 43-63)

MATETSSLKLFGINLLETTSVQNQSSEPRPGSGSGSESRYECQYCCREFANSQALGGHQ
NAHKKERQLLKRAQMLATRGLPRHNFHPHTNPLLSAFAPLPHLLSQPHPPPHMMLSPSS
SSSKWLYGEHMSSQNAVGYFHGGRGLYGGMESMAGEVKTHGGSLEPMRRFAGDSDRSSG
IKLENGIGLDLHLGLGP*

>G562 (137..1285)

ATTTGAATTTCTGGGTTTCTCTCTGTTTAAGCTTCTTCTTCTTCATCTTCTGCTTACGTT
TCTTCTTCAAGGAGCTTTTCGGATTCTTGTAGAAAGAGTCATTGTTCTCTTGAAGTGGGAAA
CCTTGAAACCATTCCTATGGGAAATAGCAGCGAGGAACCAAGCCTCCTACCAAATCAGA
TAAACCATCTTACCCCGGTGGATCAAACAAATGTTATGTCCTACCTGATTGGGCAGC
TATGCAGGCATATTATGGTCCAAGAGTAGCAATGCCTCCTTATTACAATTACGCTATGGC
TGCATCTGGTCATCTCTCTCTCTTACATGTGGAATCCTCAGCATATGATGTCACCATC
TGGAGCACCTATGCTGCTGTTTATCCTCATGGAGGAGGAGTTTACGCTCATCCCGGTAT
TCCCATGGGATCACTGCCTCAAGGTCAAAGGATCCACCTTTAACAACCTCCGGGGACGCT
TTTGAGCATCGACACTCTTAAATCTACAGGGAACACAGACAATGGATTGATGAAGAA
GCTGAAAGAGTTTGATGGGCTTGCTATGTCTCTAGGAAATGGGAATCCTGAAAATGGTGC
AGATGAACATAAAGCATCACGGAACAGCTCAGAACTGATGGTTCTACTGATGGAAGTGA
TGGGAATACAACTGGGGCAGATGAACCGAACTTAAAGAAGTCGAGAGGGAACCTCCAAC
AAAAGATGGGAAACAATTGGTTCAAGCTAGCTCATTTTCACTCTGTTTCTCCGTCAAGTGG
TGATACCGGCGTAAACTCATTCAGGATCTGGAGCTATACTCTCTCTGTTGTAAGTGC
AAATTCACACCCCTTCATGTCAATCTTTAGCCATGGTTCTCTCTGAAACTTGGCTTCA

GAACGAGAGAGAACTGAAACGGGAGCGAAGGAAACAGTCTAATAGAGAATCTGCTAGAAG
 GTCAAGATTAAGGAAACAGGCCGAGACAGAAGAACTTGCTAGGAAAGTGGAAGCCTTGAC
 AGCCGAAAACATGGCATTAAAGATCTGAACTAAACCAACTTAATGAGAAATCTGATAAACT
 AAGAGGAGCAAATGCAACCTTGTGTGGACAACTGAAATGCTCGGAACCCGAAAAGAGAGT
 CCCCAGAAATATGTTGTCTAGAGTTAAGAACTCAGGAGCTGGAGATAAGAACAAGAACCA
 AGGAGACAAATGATTCTAACTCTACAAGCAAATTCCATCAACTGCTCGATACGAAGCCTCG
 AGCTAAAGCAGTAGCTGCAGGCTGAATCGATGGTAATTCATGTCTGATTCTACTTAATTT
 GTCGACATAAACAAAGAAAATAAGTGCTACTAATTTTCAGAAAACTTGATAGATAGATAG
 TATAGTAGAGAGAGAGAGAGAGAGAGAGGTGTGATGATTATTGATCTATAAATTTTCGGA
 GAGAGAGAGGGAGAAAGAGAACTTTTCCTCCAGATGAAAATTTGGTGTATGGTTGTT
 ACTGTTAATATAGAGAGGCTTTTCTTTTATAAAATGGCTTCCTTTGTTGCA
 >G562 Amino Acid Sequence (domain in AA coordinates: 253-315)
 MGNSSSEPKPPTKSDKPSPPVDQTNVHVPDWAAMQAYYGPVAMPPIYNSAMAAAGHP
 PPPYMNWNPQHMSPSGAPYAAVYPHGGGVYAHPGIPMGSPLPQGGKDPPLTTPGTLLSIDT
 PTKSTGNTDNLMLKKLKEFDGLAMSLGNGNPENGADCHKRSRNSSETDGDSDGSDGNTTG
 ADEPKLKRSEGTPTKDGKQLVQASSFHSVSPSSGDTGVKLIQSGAILSPGVSANSNPF
 MSQSLAMVPPETWLQNERELKRERRKQSNRESARRSRLRKAETEELARKVEALTAENMA
 LRSELNQLNEKSKDLRGANATLLDKLKCSEPEKRVANMLSRVKNSGAGDKNKNQGDNDNS
 NSTSKFHQLLDTKPRAKAVAAG*
 >G591 (88..1020)
 GTAAATCTCTCTTTGAAGGTTCCCTAACTCGTTAATCGTAACTCACAGTACTCGTTTCGAG
 TCAAAGTCTCTGTCTTTAGCTCAAACCATGGCTAGTAACAACCTTCACGACAACCTTTCT
 GACCAAACCTCTCTGTATGATTCTTCGAGCAAATCCTCGGCCTTCCTAACTTCTCAGCC
 TCTTCTGCCGCCGGTTTATCTGGAGTTGACGGAGGATTAGGTGGTGGAGCACCAGCCTATG
 ATGCTGCAGTTGGGTTCCGGAGAAGAAGGAGTCACATGGGTGGCTTAGGAGGAAGTGGA
 CCAACTGGGTTTCACAATCAGATGTTTCTTTGGGGTTAAGTCTTGATCAAGGGAAGGA
 CCTGGGTTTCTTAGACCTGAAGGAGGACATGGAAGTGGGAAAAGATTCTCAGATGATGTT
 GTTGATAATCGATGTTCTTCTATGAAACCTGTTTTCCACGGGCAGCCTATGCAACAGCCA
 CCTCCATCGCCCCACATCAGCCTACTTCAATCCGTCAGGGTTCGAGCTAGGCGTGGT
 CAGGCTACTGATCCACATAGCATCGCTGAGCGGTACGTAGAGAAAGAATAGCAGAACGG
 ATCAGGGCGCTGCAGGAACCTGTACCTACTGTGAACAAGACCGATAGAGCTGCTATGATC
 GATGAGATTGTGATTATGTAAAGTTTCTCAGGCTCCAAGTCAAGGTTTTGAGCATGAAC
 CGACTTGGTGGAGCCGGTGCGTTGCTCCACTTGTACTGATATGCCTCTTTCATCATCA
 GTTGAGGATGAAACCGGGTGAGGTGGAAGGACTCCGCAACCAGCGTGGGAGAAATGGTCT
 AACGATGGGACTGAACGTCAAGTGGCTAAACTGATGGAAGAGAACGTTGGAGCCCGATG
 CAGCTTCTTCAATCAAAGGCTCTTGTATGATGCCAATCTCATTGGCAATGGCAATTTAC
 CATTCTCAACCTCCGGATACATCTTCAAGTGGTCAAGCCTGAGAACAACTCTCCACAGTAG
 GATTTCTGCAATAAAGAGTTTGTACAGCTAATCCAAGTGTCCAACATGGGTTTTTCTTCT
 GCTCTAATGACTCTGGTTTTCTTCTCTCTCTCACCAGCTTGAAAGGTAAAAAAGTGAA
 AAAGGCTTTGTAGATGGAATCAATGTAGGATTTGCAGTAGAGGGCAAAAAATGTCATAT
 AGCTCAATTGATCAAGTCTTAAAAAAAAAAAAAAAAAAAA
 >G591 Amino Acid Sequence (domain in AA coordinates: 143-240)
 MASNNPHDNLSDQTPSDDFFEQILGLPNFSASSAAGLSGVDGGLGGGAPPMMLQLGSCEE
 GSHMGGGLGGSGPTGFHNQMFPLGLSLDQKGPGFLRPEGGHGSGKRFSDDVVDNRCSSMK
 PVFHGQPMQPPPSAPHQPTSIRPRVRARRGQATDPHSIAERLRRERIAERIRALQELVP
 TVNKTDRAMIDEIVDYVKFLRLQVKVLSMNLGGAGAVAPLVTDMPLSSSVEDETGE
 RTPQPAWEKWSNDGTERQVAKLMEENVGAAMQLLQSKALCMMPISLAMAIYHSQPPDTSS
 VVKPENNPQ*
 >G8 (247..1596)
 AAAAAAAAAATATCCGTCTACTCTCTCGCCGCCGTAACATTTCCCGGCACAAAACCTTC
 TCTACTCTCACCATTCTCCATCGTAATCTCTAAATCTTCTCCATTCTCTTCTCTCC
 CGATCATCTCGAGCTCTTCGTGAGAGATTATGTGATTATGTAATCGTTGTTGCTGTAGAA
 GACGATCTCTAACAACTGATTCCTTCATCATCACCTTCGCTAGATTTGTAATTTTCAGAG
 CTTGAGATGTTGGATCTTAACCTCAACGCTGATTCTCCCGAGTCGACTCAGTACGGTGGT
 GACTCATACTTAGATCGGCAGACATCAGACAACTCCGCCGGGAATCGAGTGAAGAGTCC
 GGTACATCGACGTCGTCAATTATCAATGCCGATGGAGACGAAGACTCTTGCTCTACTCGA
 GCTTTCACTCTCAGTTTCGATATTTTAAAGTCGGAAGTAGTAGCGGCGGAGACGAAAGC

AAATATCTTCTTCTTTTTCTGTGTGAGTTGGGTTTGTAAAGTTTATCCTTTTTGTTC
TCAAAATCAAGAATCGATGGCGGAGAAGGAAGAAAAAGAACCATCGAAGTTAAATCATC
CACCGGAGTTTACGGCCAACGATTTCACTACCTCCTCGACCGTTTGGTGAAATGTTTTT
TAGCGGTGGCGTTGGATTTAGTCTTGACCAATGACTCTCGTCTCAAATTTATCTCTGA
TCCTGATGAGTTCAAGTCTTCTCTCAGCTTTTAGCTGGAGCTATGGCTTCTCCGGCGGC
AGCTGCTGTTGCCGCGCTGCTGTGGTGTCTACTGCTCATCATCAGACACCTGTGAGCTC
TGTCGGTGATGGCGGTGGAAGCGGTGGTGATGTTGACCCGAGGTTTAAGCAGAGTAGACC
AACGGGATTGATGATAACTCAACCACCGGGGATGTTTACTGTACCGCCGGGGTTAAGTCC
GGCTACTCTTTTGGATTCTCCGAGCTTCTTTGGTCTTTTTTACCTCTTCAGGGAACATT
TGGTATGACACATCAACAAGCTTTAGCACAAGTCACTGCACAAGCAGTTCAAGGCAATAA
TGTTTCATATGCAGCAATCACAACATCTGAATATCCTTCTTCTACACAACAACAACA
ACAACAACAACAAGCTTCATTGACTGAGATTCCATCATTTTCTTCTGCACCTAGGTCTCA
GATTCGAGCCTCGGTTCAAGAAACATCGCAGGGTCAGAGAGAGACTTCGGAAATATCTGT
CTTTGAGCATCGGTACAGCCTCAAAATGCTGACAAACCAGCTGATGATGGATACAACG
GCGGAAATATGGGCAGAAGCAAGTGAAGGGGAGCGATTTTCTCGGAGTTATTACAAATG
TACGCATCCAGCTTGCTCTGTCAAGAAGAAAGTGGAGAGGTCACTCGATGGACAAGTAAC
GGAAATCATCTACAAGGGTCAACACAATCATGAGCTTCTCAAAAGCGCGGTAACAATAA
CGGGAGTTGTAAAAGTTCTGATATTGCAAATCAGTTTCAACAAGTAATAGCAGTCTCAA
CAAGAGTAAGAGGGACCAGGAAACAAGCCAAGTTACAACAACAGAGCAGATGCTGAAGC
AAGTGATAGCGAGGAGGTTGGGAATGCAGAGACTAGTGTGGGAGAAAGACATGAGGATGA
GCCTGATCCCAAGCGAAGAAATACAGAAGTTCCGGGTTTTCAGAACCAGTTGCTTCATCGCA
TAGAAGTGTGACAGAGCCTAGGATTATTGTCCAAACGACGAGTGAAGTTGACCTCTTAGA
TGATGGATATAGGTGGCGCAAGTATGGTCAGAAAGTAGTCAAAGGAAATCCTTATCCGAG
GAGCTACTATAAGTGTACAAACACAGATTGCGGAGTAAGGAAACATGTAGAGAGAGCAGC
AACTGACCCAAAAGCTGTTGTAACAACATATGAAGGTAAACATAACCATGATGTTCCAGC
TGCTAGAACCAGCAGCCATCAGTTAAGACCAACAATCAACACAACACCTCAACGGTTAA
CTTCAATCATCAACAGCCTGTTGCACGTTTAAAGGCTTAAAGAAGAGCAAACTCACTTGACA
GAGAAGAAGAATACGACGGCGCTTGAGCTTTTGTGAGTTTAAATGAATCTTCTTTTGGTT
AATGAACCTGTTTTTGTGCTCAAAACACCACAGGTTTCTCTGGACAGAATCTCTGATA
TTACAGTTTCAAAGGTATGTTCTTTTATTTTCATGTTGGAATCTTCTGTGTAATCTTAAG
AAGCTTTAGGAGGTATGTAAAAAACAGATTCAAAGTTATGCCCTTATGTGAATCTTTT
TGATCATGGGATAAACAAAATTTACAGGTATCCTTTTGTCTTGTGTAATAAAAAA
AAAA

>G878 Amino Acid Sequence (domain in AA coordinates:250-305, 415-475)

MAEKKEKEPSKLSSTGVSRTISLPPRPFGEMFFSGGVGFSPGPMTLVSNLFSDFDEFK
SFSQLLAGAMASPAAAVAAAAVVATAHHQTPVSSVGDGGGSGGDVDPKQSRPTGLMI
TQPPGMFTVPPGLSPATLLDSPSFFGLFSPPLQTFGMTHQQALAQVTAQAVQGNVHMQQ
SQQSEYPSSTQQQQQQQASLTEIPSFSSAPRSQIRASVQETSQGGRETSEISVFEHRS
QPQNADKPADDGYNWRKYGQKQVKGSDFPRSYKCTHPACPVKKKVERSLDGQVTEIIYK
GQHNHELPPQKRGNNNGSKSSDIANQFQTSNSSLNKSQRDQETSQVTTTEQMSEASDSEE
VGNABTSVGERHEDEPDPKRRNTEVRVSEPVASSHRTVTEPRIIVQTTSEVDLLDDGYRW
RKYGQKVVKGNPNPYPRSYKCTTPDCGVRKHVERAATDPKAVVTTYEGKHNHDVPAARTSS
HQLRPNNQHNTSTVNFNHQPPVARLRLKEEQIT*

>G971 (131..1171)

TTTTTTTTCTTCCCTCTTTAGAACTCTCTCTCTCTCGTTTTTGACACTTATCCTCTC
TCTTTTTTCTCTCTCCCTCTCTCTCTGGCCGGAAGAAACAACGTCGTTTATAGCTAA
AGATTCGATCATGTTGGATCTTAACCTAAAGATCTTTTCTTCTTATAACGAAGATCAAGA
TCGGAAAGTACCATTAATGATCTCAACCACCGGTGAAGAAGAATCTAACTCATCTTCCTC
CTCCACAACAGACTCTGCAGCGAGAGATGCTTTCATCGCTTTTGGAATCTCAAACGCGA
CGATGACCTTGTTCCCTCCTCCTCCTCCTCATAAAGAAACAGGAGATCTCTTTCC
GGTGGTGGCTGATGCTCGTCGGAATATAGAATTTCTCCGTGGAAGACAGTCACTGGTTGAA
TCTTTCTTCTTTACAAAGAAATACACAGAAAATGGTGAAGAAGAGCAGAAGAGGACCAAG
GTCTCGTAGCTCCCAATATCGTGGCGTCACTTTTTACCGTCGCACCGGTCGTTGGGAATC
TCATATTTGGGATTGTGGAAGCAAGTTTATTTGGGCGGGTTTGATACTGCTTACGCAGC
AGCAAGGGCTTACGACCGAGCTGCTATCAAATCCCGTGGTCTCGATGCAGACATCAATTT
CGTCGTGGATGATTATAGGCATGACATCGATAAGATGAAGAATTTAAATAAGGTGGAGTT
CGTGCAAAACACTTAGGCGAGAGAGTGCAGGTTTCGGAAGAGGAAGTTCCAAATACAAAGG

CTTGGCTCTTCAAAAATGCACCCAATTCAAACTCATGATCAGATTCATCTCTTCCAAAA
 CAGGGGATGGGATGCAGCAGCAATAAAATACAATGAGTTGGGAAAGGGAGAAGGAGCCAT
 GAAGTTTGGTGCCCATATCAAAGGAAATGGTCACAATGATCTTGAACCTAAGTCTCGGAAT
 TTCATCATCATCGGAAAGTATAAAGTTGACAACAGGCGATTACTATAAGGGTATCAATCG
 GTCCACGATGGGTTTATACGGTAAGCAATCATCGATATTTTACCCATGGCAACCATGAA
 ACCTCTGAAGACAGTTGCAGCATCATCAGGATTCCTTTTATCAGCATGACAAGTTTCCTC
 TTCCTCCATGTCCAATTGTTTTGATCCATAGGATCGTTCTACACTCTCTTAACATAATATA
 TATTTTACTCTATCTGATTATTGTATACAAGGATAAAATTTGATTCTTTCCTTAATGAG
 TGAGAAATATTGGAAGTGTTAAAAA

>G971 Amino Acid Sequence (conserved domain in aa coordinates: 120-186)

MLDLNLKIFSSYNEDQDRKVLPMISTTGEESNSSSSSTTDSAADAFIAFGILKRDDDL
 VPPPPPPHKETGDLFPVADARRNIEFSVEDSHWLNLSLQRTQKMKVKSRRGPRSR
 SQYRGVTFYRRRTGRWESHIWDCGKQVYLGFDYAAAAAYDRAAIKFRGLDADINRVVD
 DYRHDIDKMKNLNKVEFVQTLRRRESASFGRGSSKYKGLALQKCTQFKTHDQIHLFQNRGW
 DAAAIKYNELGKGEGAMKFGAHIKGNHNDLELSLGISSSSSES IKLTGTDYKGINRSTM
 GLYKQSSIFLPMATMKPLKTVAASSGFPFISMTSSSSSSMSNCFDP*

>G975 (58..657)

ATTACTCATCATCAAGTTCTACTTTCTCTCTGACAAACATCACAGAGTAAGTAAGAATG
 GTACAGACGAAGAAGTTTCAAGGTTGTCAGGCAACGCCATTGGGGTTCTTGGGTCGCTGAG
 ATTCGTCATCCTCTCTTGAACCGGAGGATTTGGCTAGGGACGTTTCGAGACCGCAGAGGAG
 GCAGCAAGAGCATAACGACGAGGCCGCCGTTTTAATGAGCGGCCGCAACGCCAAAACCAAC
 TTTCCCTCAACAACAACAACACCGGAGAACTTCCGAGGGCAAAACCGATATTTAGCT
 TCGTCCACAATGTCATCCTCAACATCATCTTCATCGCTCTCTTCCATCCTCAGCGCCAAA
 CTGAGGAAATGCTGCAAGTCTCCTTCCCCATCCCTCACCTGCCTCCGTCTTGACACAGCC
 AGCTCCCATATCGGCGCTTGGCAGAAACGGGCCGGTTCAAAGTCTGACTCCAGCTGGGTC
 ATGACGGTGGAGCTAGGTCCCGCAAGCTCCTCCCAAGAGACTACTAGTAAAGCTTCACAA
 GACGCTATTCTTGTCTCCGACCTGAAGTTGAAATTTGGTGGCAGCAGAGAAGAAGTATTG
 GATGAGGAAGAAAAGGTTGCTTTGCAAATGATAGAGGAGCTTCTCAATACAACTAAATC
 TTATTTGCTTATATATATGTACCTATTTTCATTGCTGATTACAGCCAAAATAATCAATT
 ATACCGTGATTTTATAGATGTTTATATTAAAGGTTGTTAGATATA

>G975 Amino Acid Sequence (domain in AA coordinates: 4-71)

MVQTKKFRGVRQRHWGWSWVAEIRHPLLRRIWLGTFTABEAAARAYDEAAVLMSSGRNAKT
 NFPLNNNTGETSEGKTDISASSTMSSTSSSSLSILSAKLKCKCKSPSPSLTCLRLDT
 ASSHIGVWQKRAGSKSDSSWMTVELGPASSSQETTSKASQDAILAPTEVEIGGSREEV
 LDEEEKVALQMI EELLNTN*

>G994 (180..917)

TGTATATATAGTTAGTTAGTTGAGATAAACTTGGTTACCACTTTTGTGTGGTCTTTCTTT
 TTCTTTTCTCCATTTTCCATTTATCGACCCCTTGGGTGTAGCTAATTACTTTTCGCGATT
 TTCAAATCCAATAAAGTTTAAATTTGATGAAGCTTTTTTAAACCATATAATATAAATAA
 TGGGTGGTCGTAAACCATGTTGTGATGAGGTTGGATTAAGAAAGGTCATGGACAGTGG
 AAGAAGATGGGAAACTAGTTGATTTCTTAAGGGCACGTGGCAACTGCGGTGGTGGTGGAG
 GAGGATGGTGTGAGAGACGTGCCAAAACCTGGCGGGGCTAAGGAGGTGTGGCAAAAGTT
 GCCGTCTCCGGTGGACTAATTATCTCCGGCCAGATCTCAAGAGAGGTCTTTTACTGAAG
 AAGAAATCCAACCTAGTCATTGATCTTCATGCTCGCCTTGGCAATAGATGGTCAAGATTG
 CAGTGGAGTTACCAGGAAGAACAGACAACGATATCAAAAATTATTGGAACACTCATATAA
 AGAGGAAGCTTATAAGAATGGGTATTGATCCAAACACACATCGTCGATTGACCAACAAA
 AAGTCAACGAGGAGGAACGATATTGGTCAACGATCCAAAGCCTCTGTCTGAGACCGAGG
 TATCTGTTGCTTTGAAGAATGACACGTGAGCAGTGTATCAGGAAATCTAAACCAATTGG
 CTGACGTGGACGGTGATGATCAGCCGTGGAGCTTTCTAATGGAAAATGACGAAGGAGGAG
 GTGGCGACGCCCGCGGAGAGCTTACGATGCTATTGTCCGGTGACATTACGTCATCATGTT
 CTTCTTCGTATCTTTGTGGATGAAGTATGGAGAATTCGGATACGAAGATTTAGAATTG
 GATGTTTCGATGTTTAGAGATTCAAGTATGTTTAAATAGGCCGTAGGTTGATTAATCATA
 AGGTTCAATTGACTTCATTCTAGAAATGTTGATTTGGACAGTATAAAGAATCAAAGTTAT
 GAAACATTGTAATTTGATTTCCAAATTAATCTAATGAATAAATGTGCTTTGCAAAAAA
 AAAAAA

>G994 Amino Acid Sequence (domain in AA coordinates: 14-123)

MGGRKPCCDEVGLRKGPTVEEDGKLVDFLRARGNCGGGGGWCWRDVPKLAGLRRCGKS

CRLRWNTNYLRPDLKRGLFTEEEIQLVIDLHARLGNRWSKIAVELPGRDNDIKNYWNTHI
KRKLIRMGIDPNTHRRFDQQKVNNEETILVNDPKPLSETEVSVALKNDTSAVLSGNLNQL
ADVDDDDQWPWSFLMENDEGGGGDAAGELTMLLSGDITSSSSSSSLWMKYGEFGYEDLEL
GCFDV*

>G2347 (81..626)

AGCCCATCCTTCAACATTGCTTCCTAACCAGAAATCCACCATCATCTTCCCACGAATACA
ACTTAAAGCTTTTACCAGAAAATGGAGGGTCAGAGAACACAACGCCGGGGTTACTTGAAAG
ACAAGGCTACAGTCTCCAACCTTGTGAAGAAGAAATGGAGAATGGCATGGATGGAGAAG
AGGAGGATGGAGGAGACGAAGACAAAAGGAAGAAGGTGATGGAAAGAGTTAGAGGTCCTA
GCACTGACCGTGTTCATCGCGACTGTGCCAGGTGATAGGTGCACTGTTAATTTGACTG
AGGCCAAGCAGTATTACCGCAGACACAGAGTATGTGAAGTACATGCAAAGGCATCTGCTG
CGACTGTTGCAGGGGTGAGGCAACGCTTTTGTCAACAATGCAGCAGGTTTCATGAGCTAC
CAGAGTTTGATGAAGCTAAAAGAAGCTGCAGGAGGCGCTTAGCTGGACACAATGAGAGGA
GGAGGAAGATCTCTGGTGACAGTTTGGAGAAGGGTCAGGCCGGAGAGGGTTTAGCGGTC
AACTGATCCAGACTCAAGAAAGAAAACAGGGTAGACAGGAAACTTCCTATGACCAACTCAT
CATTCAAGCGACCACAGATCAGATAAACCTCCCGCTCTCTCTCTTCTGTCATCTACATA
TGCTCTATCTACACTCTTATTAGACAAATAATGGCATCTAACAATGTCAAGAAAAGTTGG
TCATGGTATTAAATCCTACACGGATATATAACTATAAACCTCTAGTCCCCCTCTATGCTGT
CTGTGAATGAATATCTATCCGGAATGTATTTCGCATAGTCTTGCGTCTAATAATGTTTAT
TGATTTTGTGTA

>G2347 Amino Acid Sequence (domain in AA coordinates: 60-136)
MEGQRTQRRGYLKD KATVSNLVEEEMENGMDGEEEDGGDEDKRKKVMERVRGPSTDRVPS
RLCQVDRCTVNLTEAKQYYRRHRVCEVHAKASAATVAGVRQRFCCQCSRPFHELPEFDEAK
RSCRRRLAGHNERRRKISGDSFGECSGRRGFSGQLIQTQERNRVDRKLPMTNSSFKRPQI
R*

>G2010 (1..525)

ATGGAGGGTAAGAGATCACAAGGACAAGGTTACATGAAAAAGAAGTCTTACCTTGTGGAA
GAAGATATGGAGACTGATACGGATGAAGAAGAGGAAGTAGGTAGGGATAGAGTTAGAGGG
TCTAGAGGTAGCATCAATCGTGGTGGCTCGTTGCGGCTTTGCCAAGTAGATAGATGCACA
GCTGATATGAAAGAGGCAAACTGTATCACCAGGAGACACAAAGTGTGTGAAGTTCATGCA
AAGGCATCTTCTGTCTTTCTCTCAGGACTTAACCAACGCTTTTGTCAACAATGCAGTAGG
TTTCATGACCTCCAAGAGTTTGATGAAGCTAAGAGAAGTTGCAGGAGGCGCTTAGCTGGA
CACAAATGAGCGAAGAAGGAAGAGCTCTGGTGAGAGTACTTATGGAGAAGGATCAGGTCGG
AGAGGAATCAATGGTCAGGTGGTGATGCAGAATCAAGAAAGATCAAGGGTAGAGATGACA
CTTCTATGCCAACTCATCATTCAAGCGACCACAGATTAGATAG

>G2010 Amino Acid Sequence (domain in AA coordinates: 53-127)
MEGKRSQGQGYMKKSYLVEEDMETDTDEEEVGRDRVRGSRGSINRGGSLRLCQVDRCT
ADMKEAKLYHRRHKVCEVHAKASSVFLSGLNQRFCCQCSRFDLQEFDEAKRSCRRRLAG
HNERRRKSSGESTYGECSGRRGINGQVVMQNQERSRVEMTLMPNPSSFKRPQIR*